

University of Kentucky

UKnowledge

---

Epidemiology and Environmental Health Faculty  
Publications

Epidemiology and Environmental Health

---

4-30-2021

## Epigenome-Wide Association Study of Kidney Function Identifies Trans-Ethnic and Ethnic-Specific Loci

Charles E. Breeze

*National Cancer Institute*

Anna Batorsky

*University of North Carolina, Chapel Hill*

Mi Kyeong Lee

*National Institute of Environmental Health Sciences*

Mindy D. Szeto

*University of Colorado Anschutz Medical Campus*

Xiaoguang Xu

*University of Manchester, UK*

Follow this and additional works at: [https://uknowledge.uky.edu/epidemiology\\_facpub](https://uknowledge.uky.edu/epidemiology_facpub)



Part of the [Biostatistics Commons](#), [Epidemiology Commons](#), [Genetics and Genomics Commons](#), and [the Medical Genetics Commons](#)

See next page for additional authors

**Right click to open a feedback form in a new tab to let us know how this document benefits you.**

---

### Repository Citation

Breeze, Charles E.; Batorsky, Anna; Lee, Mi Kyeong; Szeto, Mindy D.; Xu, Xiaoguang; McCartney, Daniel L.; Jiang, Rong; Patki, Amit; Kramer, Holly J.; Eales, James M.; Raffield, Laura; Lange, Leslie; Lange, Ethan; Durda, Peter; Liu, Yongmei; Tracy, Russ P.; Van Den Berg, David; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, TOPMed MESA Multi-Omics Working Group; Evans, Kathryn L.; Kraus, William E.; and Arnett, Donna K., "Epigenome-Wide Association Study of Kidney Function Identifies Trans-Ethnic and Ethnic-Specific Loci" (2021). *Epidemiology and Environmental Health Faculty Publications*. 75.  
[https://uknowledge.uky.edu/epidemiology\\_facpub/75](https://uknowledge.uky.edu/epidemiology_facpub/75)

This Article is brought to you for free and open access by the Epidemiology and Environmental Health at UKnowledge. It has been accepted for inclusion in Epidemiology and Environmental Health Faculty Publications by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

---

# Epigenome-Wide Association Study of Kidney Function Identifies Trans-Ethnic and Ethnic-Specific Loci

Digital Object Identifier (DOI)

<https://doi.org/10.1186/s13073-021-00877-z>

## Notes/Citation Information

Published in *Genome Medicine*, v. 13, issue 1, article no. 74.

© The Author(s) 2021

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <https://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<https://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

The first 20 authors and the author from the University of Kentucky are shown on the author list above. Please refer to the downloaded document for the complete author list.

## Authors

Charles E. Breeze; Anna Batorsky; Mi Kyeong Lee; Mindy D. Szeto; Xiaoguang Xu; Daniel L. McCartney; Rong Jiang; Amit Patki; Holly J. Kramer; James M. Eales; Laura Raffield; Leslie Lange; Ethan Lange; Peter Durda; Yongmei Liu; Russ P. Tracy; David Van Den Berg; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, TOPMed MESA Multi-Omics Working Group; Kathryn L. Evans; William E. Kraus; and Donna K. Arnett

RESEARCH

Open Access



# Epigenome-wide association study of kidney function identifies trans-ethnic and ethnic-specific loci

Charles E. Breeze<sup>1,2,3\*</sup> , Anna Batorsky<sup>4</sup>, Mi Kyeong Lee<sup>5</sup>, Mindy D. Szeto<sup>6</sup>, Xiaoguang Xu<sup>7</sup>, Daniel L. McCartney<sup>8</sup>, Rong Jiang<sup>9</sup>, Amit Patki<sup>10</sup>, Holly J. Kramer<sup>11,12</sup>, James M. Eales<sup>7</sup>, Laura Raffield<sup>13</sup>, Leslie Lange<sup>6</sup>, Ethan Lange<sup>6</sup>, Peter Durda<sup>14</sup>, Yongmei Liu<sup>15</sup>, Russ P. Tracy<sup>14,16</sup>, David Van Den Berg<sup>17</sup>, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, TOPMed MESA Multi-Omics Working Group, Kathryn L. Evans<sup>8</sup>, William E. Kraus<sup>15,18</sup>, Svati Shah<sup>15,18</sup>, Hermant K. Tiwari<sup>10</sup>, Lifang Hou<sup>19,20</sup>, Eric A. Whitset<sup>21,22</sup>, Xiao Jiang<sup>7</sup>, Fadi J. Charchar<sup>23,24,25</sup>, Andrea A. Baccarelli<sup>26</sup>, Stephen S. Rich<sup>27</sup>, Andrew P. Morris<sup>28</sup>, Marguerite R. Irvin<sup>29</sup>, Donna K. Arnett<sup>30</sup>, Elizabeth R. Hauser<sup>15,31</sup>, Jerome I. Rotter<sup>32</sup>, Adolfo Correa<sup>33</sup>, Caroline Hayward<sup>34</sup>, Steve Horvath<sup>35,36</sup>, Riccardo E. Marioni<sup>8</sup>, Maciej Tomaszewski<sup>7,37</sup>, Stephan Beck<sup>2</sup>, Sonja I. Berndt<sup>1</sup>, Stephanie J. London<sup>5</sup>, Josyf C. Mychaleckyj<sup>27</sup> and Nora Franceschini<sup>21\*</sup>

## Abstract

**Background:** DNA methylation (DNAm) is associated with gene regulation and estimated glomerular filtration rate (eGFR), a measure of kidney function. Decreased eGFR is more common among US Hispanics and African Americans. The causes for this are poorly understood. We aimed to identify trans-ethnic and ethnic-specific differentially methylated positions (DMPs) associated with eGFR using an agnostic, genome-wide approach.

**Methods:** The study included up to 5428 participants from multi-ethnic studies for discovery and 8109 participants for replication. We tested the associations between whole blood DNAm and eGFR using beta values from Illumina 450K or EPIC arrays. Ethnicity-stratified analyses were performed using linear mixed models adjusting for age, sex, smoking, and study-specific and technical variables. Summary results were meta-analyzed within and across ethnicities. Findings were assessed using integrative epigenomics methods and pathway analyses.

(Continued on next page)

\* Correspondence: [c.breeze@ucl.ac.uk](mailto:c.breeze@ucl.ac.uk); [noraf@unc.edu](mailto:noraf@unc.edu)

<sup>1</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department Health and Human Services, Bethesda, MD, USA

<sup>21</sup>Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Results:** We identified 93 DMPs associated with eGFR at an FDR of 0.05 and replicated 13 and 1 DMPs across independent samples in trans-ethnic and African American meta-analyses, respectively. The study also validated 6 previously published DMPs. Identified DMPs showed significant overlap enrichment with DNase I hypersensitive sites in kidney tissue, sites associated with the expression of proximal genes, and transcription factor motifs and pathways associated with kidney tissue and kidney development.

**Conclusions:** We uncovered trans-ethnic and ethnic-specific DMPs associated with eGFR, including DMPs enriched in regulatory elements in kidney tissue and pathways related to kidney development. These findings shed light on epigenetic mechanisms associated with kidney function, bridging the gap between population-specific eGFR-associated DNAm and tissue-specific regulatory context.

**Keywords:** Epigenetic, Kidney function, Gene regulation, Kidney development, DNA methylation

## Background

The kidney has a central role in body homeostasis through the regulation of blood pressure, fluid, and electrolytes and by removing endogenous and exogenous toxins. Reduced kidney function measured using estimated glomerular filtration rate (eGFR) defines chronic kidney disease (CKD). CKD affects 14.5% of the adult US population and is a leading cause of death and disability [1, 2]. CKD has a high burden among non-European US ethnic groups, but mechanisms for this health disparity are poorly understood [3]. A better understanding of the mechanisms influencing kidney function may provide insights into CKD occurrence and risk.

Complex interactions between genetic, lifestyle, and environmental exposures likely contribute to the observed eGFR variation across populations. DNA sequence variation accounts for 7.6% of the estimated heritability of eGFR in trans-ethnic genome-wide association studies (GWAS) [4]. Epigenetic modifications of the genome such as DNA methylation (DNAm) are heritable and contribute to gene regulation. DNAm consists of the addition of a methyl group to cytosines, typically at cytosine-guanine dinucleotides (CpG sites). DNAm is influenced by lifetime exposures and may provide clues on ethnic-specific differences influencing eGFR. Differential DNAm at CpG sites can be studied using microarrays with reasonable genome-wide coverage through epigenome-wide association studies (EWAS) [5, 6].

Recent EWAS in whole blood have identified differentially methylated positions (DMPs) associated with blood pressure and eGFR, and disease states such as CKD and rapid decline in eGFR [7–11]. Early studies had modest sample sizes (40 to 407 individuals) or were limited to a single ethnic group [7–9]. A large EWAS performed separate discovery analyses within the Atherosclerosis Risk in Communities (ARIC, 2264 African Americans) and the Framingham Heart Study (FHS, 2395 white participants) followed by cross-replication of findings [11]. The study identified 19 DMPs for eGFR or CKD. Overall, these studies support a role for DNAm in CKD-related

traits. However, previous studies did not account for important potential confounders such as smoking status and cumulative exposure in the discovery group, which have widespread effects on DNAm patterns [12] and are risk factors for CKD, nor did they assess DNAm at CpG sites across multiple ethnic groups during discovery.

The main aim of this study is to identify DNAm patterns associated with eGFR in multi-ethnic studies using data from European/European American (EA), African American (AA), and Hispanic/Latino (H/L) participants. We performed both trans-ethnic and ethnic-specific EWAS using whole blood-based Illumina DNAm data assayed in participants of the Women's Health Initiative (WHI), the Multi-Ethnic Atherosclerosis Study (MESA), and the Jackson Heart Study (JHS). We replicated our findings in the HyperGEN, Generation Scotland, and CATHGEN studies, in addition to analyzing results via look-ups in a published study [11]. We identified DMPs associated with eGFR in trans-ethnic and ethnic-specific analyses, and provided supporting evidence for the contribution of identified DMPs to kidney function and development using *in silico* approaches and human kidney tissue-specific data.

## Methods

### Study design and populations

Our study design included a discovery step comprising three population-based studies (WHI, MESA, and JHS) and three replication studies (HyperGEN, Generation Scotland, and CATHGEN) (Additional file 1: Fig. S1). WHI is a study of postmenopausal women (aged 50–79 years), comprising 161,808 women recruited from 40 US clinical centers who participated in an observational study or in clinical trials during 1993–1998 as previously described [13–16]. MESA is a multi-ethnic study of sub-clinical cardiovascular disease and risk factors for cardiovascular disease [17], consisting of 6814 asymptomatic men and women aged 45–84 (38% EA, 28% AA, 22% H/L, and 12% Asian) recruited from six field centers across the USA and examined in 2000–2002, followed by four

subsequent examination periods. JHS is a study of cardiovascular disease and its risk factors in AA, comprising 5306 African Americans aged 21 to 94 years recruited from the Jackson, MS, metropolitan area from 2000 to 2004, with four follow-up exams [18]. HyperGEN is a family-based study with a sib-pair design. Hypertensive African American sibships were recruited from Forsyth County, NC, and from the community-at-large in Birmingham, AL, from 1995 to 2000 [19]. Generation Scotland is a family-based and population-based study consisting of 23,690 European participants recruited via general medical practices across Scotland between the years 2006 and 2011 [20]. CATHGEN is a biorepository of clinical samples from a prospectively collected clinical cohort of individuals undergoing cardiac catheterization at Duke University [21]. Both discovery and replication included multi-ethnic studies. Study descriptions are shown in Additional file 2: Supplementary Methods.

### Phenotypes

Serum creatinine-based eGFR was estimated using the Chronic Kidney Disease Epidemiology equation which includes age, sex, and a constant for AA [22].

### Epigenetic data and quality control

Briefly, preprocessing included removal of probes with detection  $p$  values  $> 0.01$  in  $> 10\%$  of samples and samples with detection  $p$  values  $> 0.01$  in  $> 1\%$  of probes. Beta values were normalized using beta-mixture quantile (BMIQ) normalization method (WHI-BAA and WHI-EMPC) [23], the normal-exponential out-of-band (NOOB) preprocessing method (JHS, MESA, CATHGEN) [24], Subset Quantile Normalization (SQN) (CATHGEN) [25], or dasen (Generation Scotland) [26]. Batch effect correction was performed using ComBat [27] or adjusting batch as a covariate. CpGs overlapping with the list of potentially polymorphic sites in the relevant ethnic group and cross-reactive probes were removed [28]. Cell proportions were estimated using the reference-based Houseman method for whole blood [29]. To adjust for population structure, principal components (PCs) or ethnic-informative markers were obtained from the genome-wide genotype data available using standard methods [30]. DNAm sites were annotated to include chromosome, position, UCSC gene names, relationship to CpG islands, location in gene enhancer regions, and DNase I hypersensitive sites (DHSs) using Illumina's annotation file [31]. Detailed methods for MESA, which have not been previously published, are included in Additional file 2: Supplementary Methods.

### EWAS

Methylation betas were used as predictors for eGFR in ethnic-stratified analyses, adjusting for age, sex, smoking history (current, past, and no smoking, and pack-years),

4 to 10 principal components, cell type composition, and study-specific covariates. DMPs were modeled by fitting robust linear models (or linear mixed models in JHS to account for family relationships) and performing robust standard error calculations via the 'sandwich' package. These analyses were performed using R version 3.5.3. Because of the small samples within each ethnicity in MESA, we fitted linear models in that study without robust estimation. EWAS results were meta-analyzed across all samples and within each ethnicity using fixed-effect inverse-variance weighted methods implemented in METAL. We required a minimum of two studies for meta-analyses. We used an  $FDR < 0.05$  (Benjamini-Hochberg).

EWAS analyses were subsequently performed in HyperGEN (AA), Generation Scotland (EA), and CATHGEN (AA and EA) using the same statistical protocols as in discovery analyses. For trans-ethnic replication, we combined all the replication samples and used a Bonferroni-adjusted  $p$ -value cutoff for the 78 tests that were performed ( $p = 6.4E-04$ ). We also considered if the direction of effects between discovery and replication samples was concordant. For ethnic-specific replication, we used EA replication samples for EA or H/L discovery meta-analyses, while for AA replication, we used AA replication samples, a Bonferroni-corrected  $p < 2.1E-03$ , and consistency in direction of effects. We also attempted to replicate DMPs from a published study [11]. This published study used eGFR instead of DMPs as a predictor in models and therefore the estimates were not comparable to our study.

### In silico annotation using eFORGE and eFORGE-TF, and pathway analyses

We performed functional overlap analysis of DMPs with eFORGE version 2.0, analyzing the default top 1000 probes from EA, AA, H/L, and all-ethnic probe sets for overlap enrichment across DNase I hotspots from the Roadmap Epigenomics Consortium [32, 33]. To ascertain whether the observed enrichment was robust and associated with top probe sets below the EWAS significance threshold, we applied additional eFORGE analyses across the 5 top EA CpG sets (comprising CpGs 1–1000, 1000–2000, 2000–3000, 3000–4000, 4000–5000, ordered by  $p$ -value), to detect overlap enrichment across DNase I hotspots from the Roadmap Epigenomics Consortium [32, 33]. We thus performed integrative epigenomics analyses on data from the Roadmap Epigenomics consortium [34] using the eFORGE framework (<https://eforge.altiusinstitute.org/>) [32, 33]. To further understand eFORGE enrichment results, we performed TF motif analysis on the probes underlying eFORGE tissue-specific enrichment signal for kidney. We used the eFORGE-TF module, seeking to uncover

the main TF motifs associated with our DNase I hotspot enrichment for top sites [33]. We used PANTHER analysis via the AmiGO framework to uncover pathways associated with the identified TF motifs [35].

#### meQTL data in the whole blood

We used publicly available data from the biobank-based integrative omics study (BIOS) QTL database [36, 37] and cis-meQTL data from the FHS ( $n = 4170$  participants, 450K DNAm panel, 1000G imputed data, defined within 2-Mb window) [38]. We also used meQTL data from mQTLdb [39].

#### meQTL in normal kidney tissue

We analyzed a total of 211 individuals with matching kidney genome-epigenome information from TRAN Scriptome of renal humAn TissuE Study (TRANSLATE), an extension of the TRANSLATE study (TRANSLATE-T), Renal gEne expreSsion and PredisPOsition to cardiovascular and kidNeY Disease (RESPOND), and molecular analysis of mechanisms regulating gene expression in post-ischemic injury to renal allograft (REPAIR) ( $n = 192$ ), in addition to normal samples with available genotype and kidney DNAm profiles from the National Institutes of Health (NIH) Tissue Cancer Genome Atlas (TCGA) ( $n = 19$ ) [40–43]. Kidney tissue samples from TRANSLATE, RESPOND, and TCGA were taken from the healthy, unaffected by cancer part of the organ after elective nephrectomies, and renal specimens in TRANSLATE-T and REPAIR studies were collected as pre-implantation biopsies from deceased donors' kidneys before transplantation [40, 42, 43]. Information on local recruitment teams, genotyping, and DNAm methods are included in Additional file 2: Supplementary Methods. In total, 374,826 CpG sites were available for further analyses after quality control filters. For post-EWAS DMP analysis, we conducted kidney cis-meQTL analysis on 62 DMPs (out of 374,826) from the study. For GWAS SNP analysis, we conducted kidney cis-meQTL analysis on a set of published eGFR GWAS SNPs [44]. The cis-meQTL analysis was conducted on 195 kidney DNA samples that passed all quality control criteria. For analysis, we used the FastQTL pipeline [45]. We used normalized  $M$ -values and genotype information for all genotyped and imputed variants passing the quality control filters under an additive mode of inheritance. Regression models included age, sex, genotyping array, source of tissue indicator (nephrectomy/kidney biopsy), the top three PCs derived from genotyped autosomal variants (genotype PCs), and six PCs derived from methylation array control probes (methylation PCs). The FastQTL cis-region was defined as  $\pm 1$  Mb from each tested CpG/SNP position. For EWAS DMP blood-kidney meQTL comparison analysis, we compared

DMP+kidney meQTL CpGs to blood meQTL CpGs from mQTLdb and calculated the percentage of overlap for data from both sources.

#### eFORGE analysis for kidney meQTL CpGs

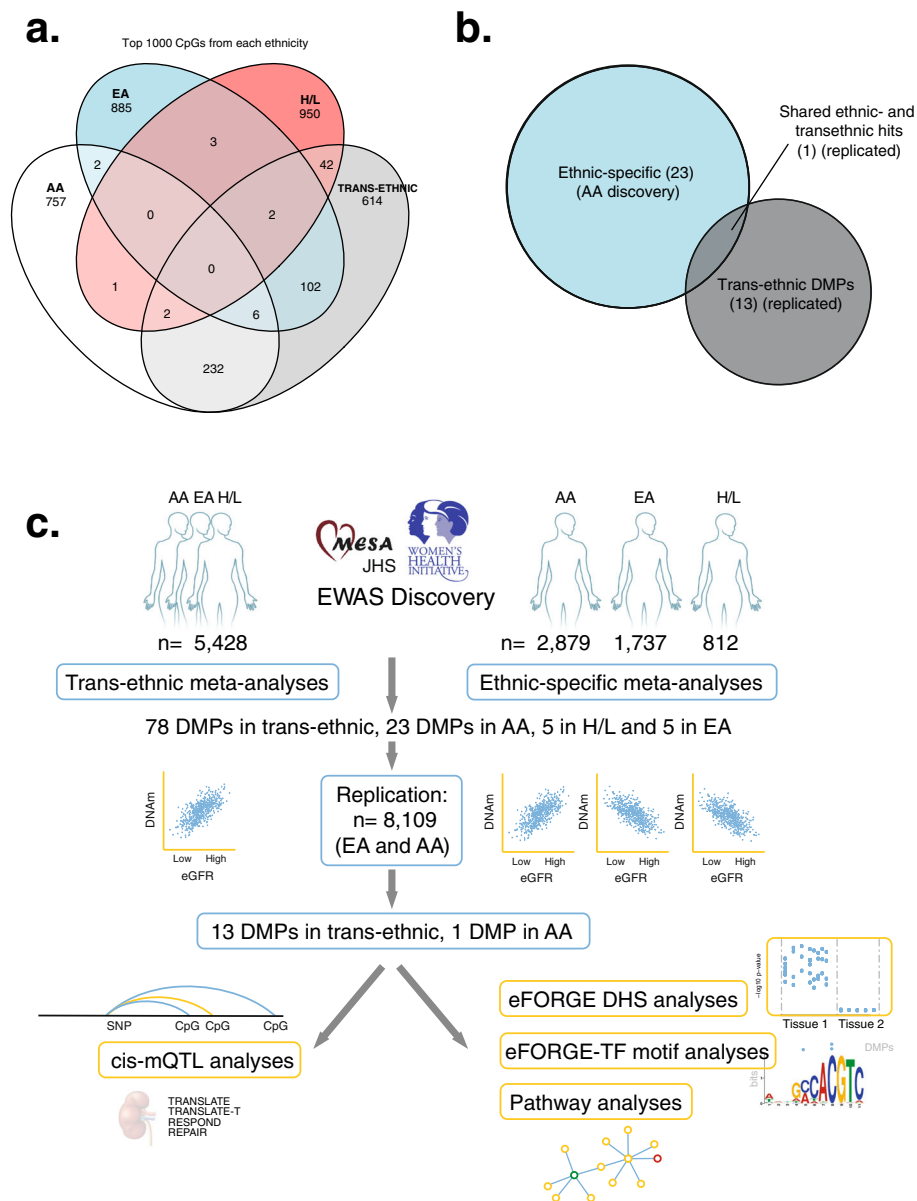
Standard eFORGE analyses were performed using default settings. For kidney meQTL CpGs, we analyzed eGFR SNPs overlapping kidney DNase I hotspots (top category from <https://forge2.altiusinstitute.org/files/0x454A546E8E9C11EA8198793F5BFE3F98/index.html>). We searched for these SNPs in the kidney meQTL file, obtaining preliminary associated meQTL CpGs (nominal  $p$ -value  $< 0.05$ ) and preliminary non-associated meQTL CpGs (nominal  $p$ -value  $> 0.95$ ), for the same set of SNPs. Standard eFORGE analyses were performed on both of these sets.

## Results

### Overview of EWAS results

The study design is shown in Fig. 1 and Additional file 1: Fig. S1. Discovery EWAS meta-analyses included up to 5428 individuals, with 2879 AA, 1737 EA, and 812 H/L participants (Additional file 3: Table S1). Quality control of DNAm data and protocol analyses for each study are shown in Additional file 3: Table S2. We performed analyses within each study and ethnic group using standard statistical protocols followed by meta-analyses of the overall samples and within each ethnicity. The quantile-quantile and Manhattan plots for meta-analyses are shown in Additional file 1: Fig. S2. Lambdas for meta-analyses were 0.987 (AA), 1.001 (EA), 1.065 (H/L), and 1.194 (trans-ethnic meta-analysis). Across our discovery analyses, we identified a total of 93 DMPs associated with eGFR at an FDR of 0.05. Of these, 78 DMPs were identified in trans-ethnic meta-analyses, 23 DMPs in the meta-analysis of AA, 5 in the meta-analysis of H/L, and 5 in the meta-analysis of EA, with some overlap in DMP findings between trans-ethnic and ethnic-specific results.

Indeed, most of the DMPs identified in trans-ethnic analyses were also present in one or more ethnic groups (Fig. 1a). However, the overlap between DMPs of AA and EA was small. Trans-ethnic replication included up to 8109 participants from three studies (Generation Scotland, CATHGEN, and HyperGEN), composed of participants of EA and AA (9% of replication samples) (Additional file 1: Fig. S1) [11]. Replication of EA or H/L findings included Generation Scotland and CATHGEN EA samples ( $n = 7349$ ) and AA included CATHGEN and HyperGEN samples ( $n = 760$ ). Among the significantly identified trans-ethnic DMPs (Bonferroni corrected  $p$ -value and consistent direction of effects), 13 of 78 replicated (Table 1), and among the significantly identified DMPs in ethnic-specific meta-analyses, 1 of 5 AA DMPs replicated (cg14871770 at *CYP2C9*) (Additional file 3:



**Fig. 1** Overview of trans-ethnic and ethnic-specific CpGs associated with kidney function. **a** Venn diagram showing trans-ethnic and unique CpGs across the top 1000 sites for European Americans (EA), African Americans (AA), Hispanic/Latino (H/L), and trans-ethnic groups. **b** Euler diagram showing the number of overlapping CpGs (1) between trans-ethnic replicated DMPs (13) and discovery ethnic-specific DMPs for African Americans -AA- (5). **c** Study design, consortium information, sample size, and the number of significant DMPs for both trans-ethnic and ethnic-specific EWAS analyses. Details shown both for discovery (top) and replication analyses (bottom). For these analyses, consortia include the Women's Health Initiative (WHI), the Jackson Heart Study (JHS), MESA, HyperGEN, Generation Scotland, and CATHGEN. In addition, we used kidney DNAm data from the TRANSLATE, TRANSLATE-T, RESPOND, and REPAIR studies for *cis*-meQTL analyses and Roadmap Epigenomics data for eFORGE DHS analyses

Table S3). Despite independent replication in AA, cg14871770 overlapped with trans-ethnic replicated DMPs (Fig. 1b). Twelve additional DMPs had replication (with consistent direction of effects) in trans-ethnic meta-analyses (Additional file 3: Table S3a). The DMP cg14871770 overlapped between trans-ethnic and AA

meta-analyses, and the cg17944885 at the *ZNF20-ZNF788P* locus was previously described [11]. Several of the replicated DMPs were expression quantitative trait methylation loci (eQTM) or *cis*-meQTL CpGs in whole blood in FHS and (BIOS QTL) (Table 1) [36–38]. Replication results for all DMPs are shown in Additional file 3:

**Table 1** Main findings from trans-ethnic EWAS meta-analyses for 13 replicated DMPs

DMP	Chr	Position (hg38)	Gene	Our study discovery		Sample size by DNAm array		Replication GS/CATHGEN/HyperGEN		Combined discovery and replication	
				Effect	p	450K	EPIC 850 K	Effect	p	Effect	p
cg13235761 <sup>^</sup>	1	203,592,452		-37.61	4.33E-06	3048	2378	-25.28	5.08E-05	-29.81	1.88E-09
cg26099045 <sup>^*</sup>	2	64,064,666		14.82	5.74E-06	3050	2378	11.24	1.01E-04	12.81	3.26E-09
cg04428662 <sup>^Ω*</sup>	4	2,932,461	<i>MFSD10</i>	-30.84	3.21E-06	3049	2378	-34.94	3.43E-07	-32.82	5.51E-12
cg23174201 <sup>^*</sup>	5	151,674,695	<i>SPARC</i>	-35.83	5.10E-06	3050	2378	-34.87	1.72E-07	-35.27	4.02E-12
cg17170437	6	44,229,461	<i>SLC29A1</i>	-47.62	1.96E-06	N/A	2378	-18.62	1.49E-04	-24.24	3.80E-08
cg14871770	10	96,658,622	<i>CYP2C9, CYP3A1</i>	-53.12	1.98E-06	N/A	2378	-25.87	1.70E-04	-33.36	1.23E-08
cg02157636	11	68,709,367	<i>TESMIN</i>	-53.21	2.15E-06	N/A	2378	-27.25	1.14E-05	-33.33	8.58E-10
cg26039141 <sup>^Ω</sup>	11	75,402,116	<i>RPS3</i>	-39.14	3.40E-06	3050	2378	-33.42	1.78E-05	-36.06	2.90E-10
cg22593432 <sup>^*</sup>	13	32,001,768		-29.91	1.54E-07	3050	2378	-17.66	1.79E-04	-22.64	4.61E-10
cg11789371 <sup>^*</sup>	14	102,085,048	<i>HSP90AA1</i>	-35.40	4.44E-06	3050	2378	-22.66	3.58E-04	-27.80	1.41E-08
cg05796561	18	57,128,273		-45.24	2.96E-06	N/A	2378	-22.91	1.69E-04	-29.24	1.41E-08
cg17944885 <sup>^Ω*</sup>	19	12,114,920	<i>ZNF20, ZNF788P</i>	-32.71	1.41E-09	3048	2378	-16.34	5.67E-07	-20.72	1.24E-13
cg15787712 <sup>^</sup>	19	13,837,429	<i>LOC284454, MIR23</i>	-37.60	6.00E-09	3050	2378	-28.65	2.06E-05	-33.30	9.01E-13

p-value for replication < 6.4E-04 (Bonferroni correction for 78 DMPs tested). Four of the 13 DMPs that replicated were non-450K probes (cg17170437, cg14871770, cg02157636, cg05796561). Results from combined discovery and replication are also shown

GS Generation Scotland, N/A not applicable

<sup>^</sup>cis-meQTL from the BIOS QTL [36, 37]

<sup>Ω</sup>Expression quantitative trait methylation (eQTM) data from the BIOS QTL (enrichment p-value < 0.01) [36, 37]

<sup>\*</sup>cis-meQTL in the Framingham Heart Study [38]

Tables S3a (trans-ethnic analyses) and Additional file 3: Table S3b (ethnic-specific analyses). Forest plots for each of the replicated DMPs are shown in Additional file 1: Fig. S3. We also replicated 6 DMPs identified in a prior publication (DMPs located at genes *DAZAP1*, *KIAA1549L*, *TUBGCP4/ZSCAN29*, *JAZF1*, *ZNF20-ZNF788P*, *LDB2*) (Additional file 3: Table S4) [11].

#### Overlap of eGFR-associated DMPs with genes and regulatory elements

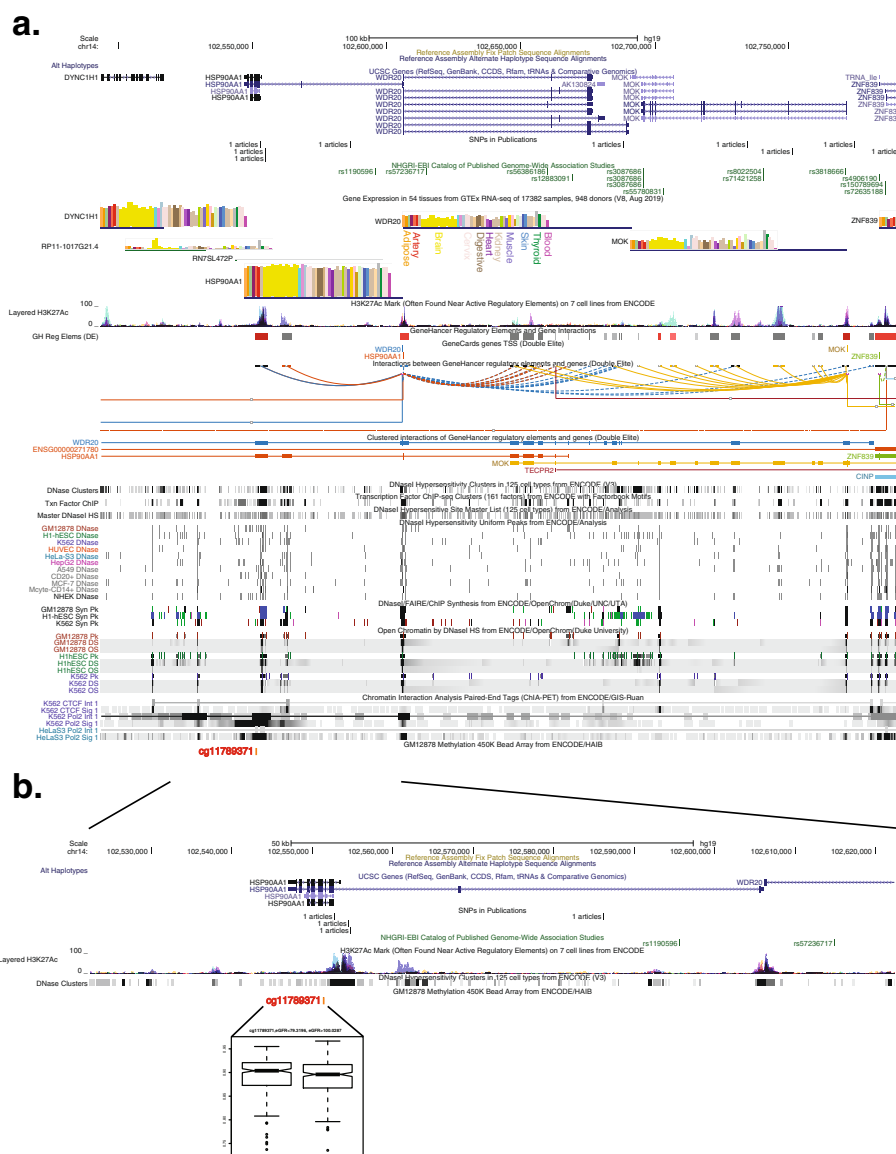
To understand the regulatory context of our DMPs, we annotated the 13 replicated significant DMPs with the closest gene and other information including epigenomic peaks, tissue-specific gene expression via RNA-seq, and chromatin interaction annotations. One of our DMPs (cg11789371) is located in an intron of *HSP90AA1*, a gene involved in protecting kidney tissue from inflammation, ischemia, and oxidative damage, and assisting with cellular repair (Fig. 2) [46]. HSP90 (heat shock protein 90) has a physiological role in eGFR regulation through the nitric oxide pathway and is a drug target candidate for kidney diseases [46]. Indeed, inhibition of HSP90 has been shown to reduce eGFR in animal models [46]. DMP cg11789371 overlaps epigenomic annotations from the ENCODE consortium including DNase I hypersensitive sites in kidney cells, among other cell types. This DMP is a cis-meQTL CpG regulated by rs11621083, a variant located upstream in an intron of *HSP90AA1*. The DMP also forms part of a GeneHancer

interacting site, contacting the promoter of *WDR20* and an alternative promoter of *HSP90AA1* which is located 50 kb away. The entire region surrounding this DMP contains a number of genes expressed in the kidney, several of which seem to interact with these two promoters. Genes from this locus presenting RNA-seq expression in the kidney include *DYNCCI1*, *MOK*, *ZNF839*, *WDR20*, and *HSP90AA1*. Taken together, annotations for our DMP cg11789371, which overlaps an intron of *HSP90AA1*, a gene associated with eGFR regulation, suggest a potential link between DNA methylation and eGFR regulation through the nitric oxide pathway.

#### DMPs for eGFR are enriched for kidney regulatory function related to kidney development

To further understand the regulatory potential and chromatin context of our EWAS findings across different tissues, we performed integrative epigenomics analyses on data from the Roadmap Epigenomics consortium [34] using the eFORGE framework (<https://eforge.altiusinstitute.org/>) [32, 33]. We found enrichment for kidney-specific DNase I hotspots, which has also been described in GWAS of eGFR (Fig. 3a, d) [4, 47]. eFORGE showed consistent enrichment results for kidney-specific DNase I hotspots when applied to the top EA discovery probes, showing a corresponding trend with study p-value (analyses performed using the "EPIC" setting with 1000 repetitions, BY correction). In addition, analysis with eFORGE-TF uncovered significant enrichment for several





**Fig. 2** eGFR-associated differentially methylated position cg11789371. **a** *HSP90AA1* gene browser shot showing (from top to bottom) genome coordinates, local genes, NHGRI/EBI GWAS catalog SNPs, GTEx gene expression quantified via RNA-seq across different tissues, H3K27ac peaks across 7 ENCODE cell lines, GeneHancer regulatory elements, Genecards TSSs, GeneHancer chromatin interactions, ENCODE chromatin accessibility and chromatin interaction tracks, and location for eGFR-associated DMP cg11789371. **b** Expanded browser shot showing genome coordinates, local genes, NHGRI/EBI GWAS catalog SNPs, H3K27ac peaks across 7 ENCODE cell lines, and a boxplot indicating DNAm values at cg11789371 for bottom and top quartiles of eGFR, respectively. These data indicate our DMP overlaps an intron of *HSP90AA1*, a gene expressed in kidney tissue, and a DHS from ENCODE, which was detected in kidney tissue. Our DMP is also proximal to an H3K27ac peak, an RNA Polymerase 2 region determined by ENCODE ChIA-PET across several cell lines, and the promoter of *HSP90AA1*. All browser shots were generated using the UCSC genome browser (<https://genome.ucsc.edu/>) on human genome build hg19

transcription factor (TF) motifs (including motifs for *OSR1*, *OSR2*, *TBX1*, and *PAX2*). *OSR1*, *OSR2*, *TBX1*, and *PAX2* have all been shown to have roles in kidney development (Fig. 3c) [48–50]. To further understand the pathways underlying TF motif enrichment, we performed PANTHER pathway analysis using significant TF motifs [35]. This analysis uncovered pathways associated with

kidney development, including metanephros development ( $8.1 \times 10^{-3}$ ), mesonephros development ( $9.7 \times 10^{-3}$ ), and retinoic acid receptor signaling pathway ( $6.4 \times 10^{-3}$ ) (Additional file 3: Table S5). Overall, these findings suggest that epigenetic changes related to eGFR are enriched in kidney regulatory regions and pathways related to kidney development.

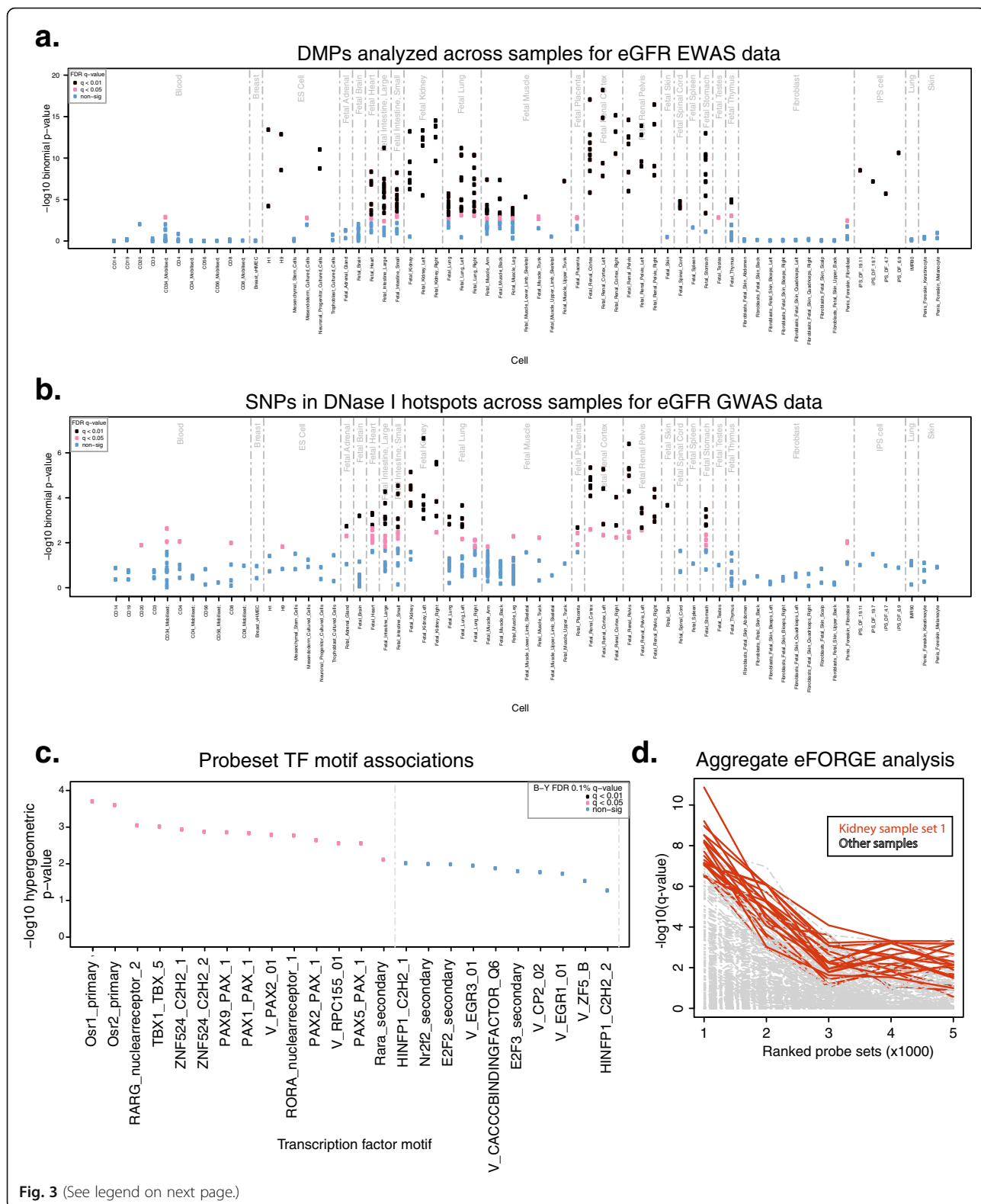


Fig. 3 (See legend on next page.)

(See figure on previous page.)

**Fig. 3** Tissue-specific integrative analysis indicates potential effect on kidney and relation with eGFR GWAS loci. **a** eFORGE analysis for top 1000 eGFR CpGs: the *x* axis indicates tissues/cell type samples used in the analysis; the *y* axis shows eFORGE enrichment ( $-\log_{10} p$ -value) of the CpG set with DNase I hotspots for a range of tissue samples (significant samples in black). The highest ranked sample set (highest black points) shows the most significant enrichment is for kidney samples, which are highly ranked for the top 1000 CpGs associated with eGFR. **b** FORGE2 analysis for eGFR SNPs from GWAS catalog: the *x* axis indicates tissues/cell type samples used in the analysis; the *y* axis shows FORGE2 enrichment ( $-\log_{10} p$ -value) of the SNP set with DNase I hotspots for a range of tissue samples (significant samples in black). The highest ranked sample set (highest black points) shows the most significant enrichment also is for kidney samples, which are highly ranked for the top 249 SNPs associated with eGFR (taken from the GWAS catalog, <https://www.ebi.ac.uk/gwas/>, downloaded 10 April 2020). **c** TF motif enrichment results for EA probes driving eFORGE tissue-specific enrichment signal: the *x* axis indicates TF motifs from TRANSFAC, JASPAR, Taipale/SELEX, and Uniprobe databases; the *y* axis shows eFORGE-TF enrichment ( $-\log_{10}$  hypergeometric *p*-value) of the input DMP set with TF motifs overlapping open chromatin sites for fetal kidney samples. Enrichment values for each TF motif are colored according to BY FDR-corrected *q*-value. A number of TF motifs involved in kidney development overlap top EA probes including *OSR1*, *OSR2*, *TBX1*, and *PAX2*. **d** Aggregated eFORGE results for EA probes: the *x* axis indicates sets of the top ranked DMPs used in the analysis (each set contains 1000 DMPs); the *y* axis shows eFORGE enrichment ( $-\log_{10} p$ -value) of each of the DMP sets with open chromatin sites for kidney (red) and other tissue samples (gray). The highest ranked probe set (set 1, left) shows the most significant enrichment for kidney samples, which remain highly ranked for probe sets 2–5, in decreasing order of study *p*-value

### eGFR GWAS variants, meQTL CpGs, and tissue-specific DNase I hotspots

Previous reports and our analyses confirm that eGFR GWAS variants are enriched for kidney DNase I hotspots (Fig. 3b, Additional file 1: Fig. S4) [4, 47], indicating that genotypes might function through kidney-specific regulatory pathways, potentially converging with identified EWAS DMPs through mechanisms which are not fully understood. To explore this further, we used mQTLdb [39] to identify meQTL CpG targets of eGFR-associated GWAS SNPs obtained from the GWAS catalog [44, 51]. These significant meQTL CpGs linked to GWAS SNPs revealed significant overlap with eGFR EWAS sites in our study ( $p < 0.002$ , Fig. 4a) and presented significant enrichment for kidney, renal cortex, and renal pelvis, among other tissues (Fig. 4b, d, Additional file 1: Fig. S5). These results support a model in which meQTL CpGs linked to eGFR GWAS SNPs overlap with EWAS DMPs and tend to localize to kidney DNase I hotspots, with potential involvement in their regulatory action (Fig. 4c). To assess these findings in human kidney tissue, we analyzed meQTL CpGs of eGFR SNPs localizing to kidney DNase I hotspots. The meQTL CpGs were identified from DNAm data from 195 normal kidney tissue samples (acquired from elective nephrectomies—taking the non-cancer affected segment—or from kidney donors) as reported previously [40–42]. While kidney meQTL CpGs not associated with eGFR SNPs (nominal  $p$ -value  $> 0.95$ ) showed no enrichment in kidney DNase I hotspots (Additional file 1: Fig. S6), the kidney meQTL CpGs with a nominal  $p$ -value  $< 0.05$  were enriched in kidney DNase I hotspots (Additional file 1: Fig. S7) for the same GWAS SNPs. These results support the aforementioned findings using eGFR SNP-associated meQTL CpGs from mQTLdb [39]. Additionally, we evaluated the overlap between kidney meQTL CpGs and blood meQTL CpGs (mQTLdb) for our top EWAS DMPs. The majority (58.3%) of these kidney meQTL CpGs were also mQTL CpGs in blood.

## Discussion

### Main findings

This study used trans-ethnic and ethnic-specific analyses of multi-ethnic cohorts to identify and replicate DMPs associated with eGFR. Our main findings include replicated DMPs at 13 sites from trans-ethnic analyses and 1 DMP from AA-specific analyses, which may reflect the larger AA discovery sample compared to other ethnic groups. All associations were newly identified in this study, except for the *ZNF20-ZNF788P* locus, which was previously described in two separate eGFR EWAS, including a population-based cohort and an HIV-infected cohort [11, 52]. Several DMPs were associated with accessible chromatin sites in kidney tissue, suggesting a regulatory role for these sites. Overall, our study identified 12 previously unreported DMPs that replicated, in addition to 6 previously published DMPs for eGFR among our 93 discovery DMPs [11]. These findings support an association between DNAm and eGFR, reflecting a convergence of lifetime influences of genetic effects, lifestyle, behaviors, and environmental exposures [53–55].

### Differences between ethnicities

Our approach of studying multi-ethnic groups identified DMPs from trans-ethnic and ethnic-specific analyses. This approach contrasts with a prior study that performed separate EA and AA analyses with cross-replication across these groups [11]. Across our two largest ethnic-specific samples (AA and EA), there was little overlap between DMPs from discovery analysis, although several discovery ethnic-identified DMPs did overlap with trans-ethnic findings (Fig. 1a). The DMP cg14871770, which is located between *CYP2C9* and *CYP2C19* (cytochrome P450 family 2 subfamily C members 9 and 19), was identified both in ethnic-specific and trans-ethnic analysis. The closest GWAS SNP to this DMP is rs4110517, a SNP associated with blood pressure identified in a multi-ethnic cohort [56]. *CYP2C9* and *CYP2C19* encode members of the cytochrome P450 superfamily of enzymes. Cytochrome P450

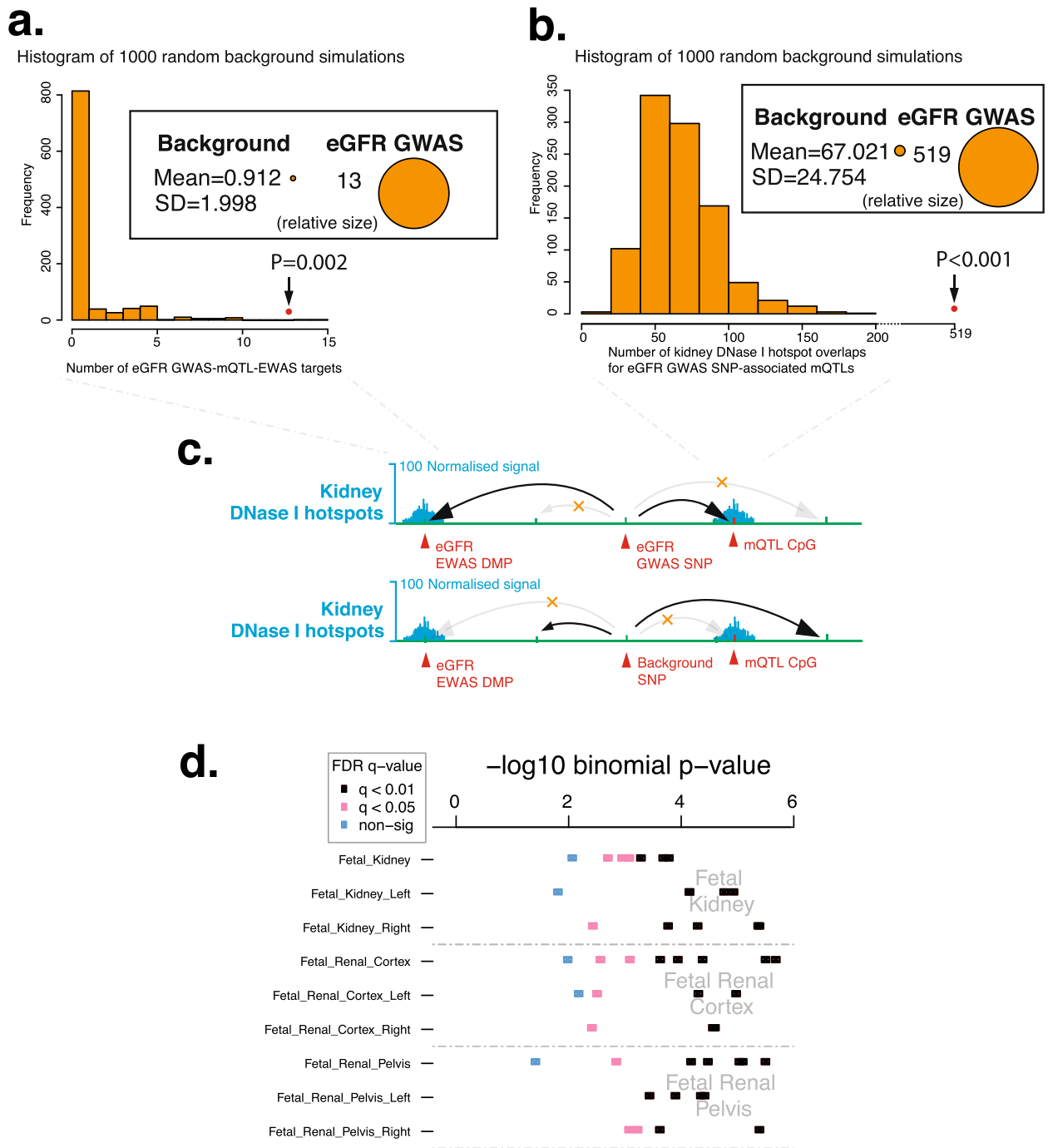


Fig. 4 (See legend on next page.)

(See figure on previous page.)

**Fig. 4** eGFR EWAS CpGs present a significant overlap with eGFR GWAS-driven meQTL effects. **a** Histogram of 1000 random background simulations (249 random SNPs each), for EWAS-meQTL overlap across the ARIES blood meQTL dataset (<http://www.mqtl.org/>). Two hundred forty-nine unique significant SNPs from the eGFR GWAS by Hellwege et al. yield 13 SNP-meQTL-EWAS DMP sites in the Aries cohort ( $p = 2.0E-03$ , empirical test, red dot and arrow), while background SNP sets overlap a mean of 0.912 SNP-meQTL-EWAS sites. **b** Histogram of 1000 random background simulations (249 random SNPs each), for meQTL-kidney DNase I hotspot overlap across Roadmap Epigenomics “Kidney” sample datasets ([https://egg2.wustl.edu/roadmap/web\\_portal/](https://egg2.wustl.edu/roadmap/web_portal/)). Two thousand seven hundred thirty-three meQTL targets of 249 unique significant SNPs from the eGFR GWAS by Hellwege et al. overlap Roadmap kidney DNase I hotspots 519 times ( $p < 0.001$ , empirical test, red dot and arrow), while background SNP sets overlap Roadmap kidney DNase I hotspots a mean of 67.021 times ( $SD = 24.754$ ). **c** Schematic showing the association of eGFR GWAS SNPs with meQTL target CpGs and eGFR EWAS CpGs (both in red text), some of which overlap kidney-specific DNase I hotspots (shown in blue, arrows indicate statistical association—not genomic contact). For comparison, a representation of a background SNP is shown. **d** Results from eFORGE analysis of significant ARIES meQTL CpGs associated with eGFR GWAS SNPs, indicating a higher-than expected overlap with the kidney, renal cortex, and renal pelvis DNase-seq hotspots (for additional results, see Additional file 1: Fig. S5)

proteins are monooxygenases that catalyze many reactions involved in the synthesis of cholesterol, other steroids, and other lipids, and in drug metabolism. *CYP2C19* genotype is associated with the metabolism of compounds influencing both renal function and hypertension [57]. The relevance of ethnic-specific findings to clinical phenotypes will need further evaluation in studies with larger samples inclusive of multiple ethnicities to define improved DNAm signatures for eGFR that may be unique to one single ancestry. Our findings suggest reduced utility of DNAm biomarkers for eGFR in diverse populations if discovery EWAS is performed in a single homogenous population.

#### Relevance of *HSP90AA1* locus

Among other findings, we report a replicated trans-ethnic DMP (cg11789371) associated with eGFR that localizes to an intron of *HSP90AA1* (heat shock protein 90 alpha family class A member 1), a gene expressed in podocytes, parietal epithelial cells, proximal tubular cells, endothelium, and mesangial cells in normal kidney tissue, with gene expression increasing in glomerulonephritis and acute kidney injury [46]. The gene product, HSP90, regulates renal blood flow and eGFR through nitric oxide metabolism and plays a role in protein folding [46]. HSP90 and other heat shock proteins are candidate drug targets for a variety of kidney diseases [46, 58]. Treatment with radicicol, an inhibitor of HSP90, has been shown to reduce eGFR in animal models [58]. DMP cg11789371 overlaps an accessible chromatin region in kidney cells, and a GeneHancer interacting site contacting the promoters of *WDR20* and *HSP90AA1*. These and other different annotations point to a potential regulatory role in kidney tissue for this eGFR-associated DMP.

#### Integrative epigenomics and pathway analysis

We detected a significant overlap of eGFR GWAS SNP-associated meQTL CpGs with kidney DNase I hotspots.

Additionally, our eGFR EWAS DMPs were enriched for kidney DNase I hotspots. These findings suggest potential links between the regulatory action of both genotypes and epigenetic DNAm elements. Importantly, these integrative epigenomic analyses considered all tissues available instead of only kidney tissue.

While some regions of the methylome show tissue specificity [59], EWAS have also shown that some DMPs are shared across different tissues, e.g., the *AHRR* locus DNAm patterns in response to smoking are shared across multiple tissues [60]. Pan-tissue findings for *AHRR* suggest similar underlying pathways in response to the same environmental stimulus [60]. Regarding discrepancies in trans-ethnic results, both genetic and environmental differences could be at play, potentially interacting with each other. In this context, our findings of both a kidney-specific DNase I hotspot and GWAS meQTL enrichment for a whole blood-based EWAS could be due to both genetic and environmental origins and warrant further research of kidney tissue DNAm in association with eGFR. Indeed, such results raise the intriguing hypothesis that tissue-specific enrichments observed separately in GWAS and EWAS might be related by the same genomic variants, thus aiding the integration of both approaches (Fig. 4a). It is important to highlight that both whole blood and kidney tissue eQTLs were obtained from individuals of European ethnicity [40–43].

The identification of eGFR DNAm signature-associated pathways is an important step towards characterizing epigenetic mechanisms for this physiologic trait and may provide clues to underlying mechanisms for CKD. Identified DMPs highlight an association with pathways of kidney development, which can influence nephron endowment at birth and subsequent CKD risk [48–50, 61]. Our *in silico* results were influenced by our DNAm findings in healthy adult kidney. Therefore, pathway results from this study support epigenetic effects during developmental windows with long-term influence on eGFR, which warrant further investigation.

### Limitations

This study is limited by the use of whole blood as the main tissue (chosen due to its availability). The discovery datasets included both the Illumina 450K and the EPIC 850K arrays, which contributed to differences in sample sizes and power to detect associations for some CpGs (Table 1). Post hoc power analyses suggest adequate power to detect DNAm differences of the range observed in the study (Additional file 2: Supplementary Methods). The methods for normalization of the DNAm beta values and study-specific quality control varied (Additional file 3: Table S2). However, we applied standardized protocols for data harmonization and statistical analyses in addition to stringent quality control as part of our meta-analyses. Our findings showed no heterogeneity of effects across studies (Additional file 1: Fig. S3), suggesting that results are robust to study-specific quality control and normalization procedures. Additionally, our reported DMPs replicated in independent samples, further validating our results. It is important to consider whether these DNAm sites associated with eGFR in blood are also applicable to effects in kidney tissue. We attempted to answer this by examining our identified meQTLs in normal kidney tissue in the TRANSLATE study. However, kidney tissue studies still have small sample sizes and lack ethnic diversity. While epigenomic mapping consortia such as Roadmap Epigenomics and ENCODE have made important steps to increase the free availability of a wide range of tissue and cell type-specific datasets, the important issue of including additional epigenomics mapping data sets for other ancestries remains to be addressed. It is important to highlight that we only observe eFORGE kidney enrichment for top EA probes. While this kidney-specific DNase I hotspot enrichment has been further confirmed by analyzing ranked DMPs in study *p*-value order (Fig. 3d), it is not apparent for DMPs from other ethnicities, or for DMPs for analysis comprising all ethnicities. DNase-seq datasets for these samples originate from the Roadmap Epigenomics consortium [34], which focused mainly on tissue samples obtained from EA individuals. Without datasets from diverse ethnic groups, it will be difficult to conclusively study inter-ethnic epigenomic variability or perform tissue-specific analyses for loci from GWAS and EWAS performed on individuals of non-European origin.

### Conclusions

We identified trans-ethnic and ethnic-specific differential DNAm positions, validated prior published

associations, and showed that several eGFR DMPs identified in this study replicated in independent samples. We have also shown that some of the DMPs are meQTL CpGs, many of which are associated with pathways relevant for kidney tissue regulation and development. Our findings include a DMP at *HSP90AA1*, a gene involved in the regulation of eGFR in kidney tissue. Identification of trans-ethnic and ethnic-specific DMPs and elucidation of their potential functional impact are preliminary steps towards identifying disease-associated epigenetic mechanisms that are specific to a particular population or shared across different populations.

### Supplementary Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1186/s13073-021-00877-z>.

**Additional file 1: Supplementary Figures. Figure S1 to S7.** Study design, QQ plots, Manhattan plots, forest plots and eFORGE/FORGE2 analyses.

**Additional file 2: Supplementary Methods.** Description of different constituent cohorts and studies, in addition to power analyses.

**Additional file 3: Supplementary Tables. Tables S1 to S5.** Descriptive characteristics, quality control and processing, and results.

### Acknowledgements

The authors thank the CATHGEN Steering Committee, the faculty and staff of the Duke cardiac catheterization lab, and the CATHGEN participants for making this work possible. We are grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team that includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants, and nurses. The authors also wish to thank the staff and participants of the JHS. The authors thank the WHI investigators and staff for their dedication and the study participants for making the program possible.

### Authors' contributions

Advised or performed statistical analyses (CEB, AB, MKL, JCM, MDS, RJ, DLM, XX, AP, JME, APM, REM, SB, SIB, SJL). Study design (CEB, SJL, NF). Generated data (YL, RPT, DVDB, EAW, LH, HJK, LR, LL, EL, PD, KE, WEK, SS, HKT, XJ, FJC, AAB, SSR, MI, DKA, ERH, JIR, AC, CH, SH, MT). CEB and NF wrote the paper with contributions from all other authors. All authors read and approved the final manuscript.

### Funding

This study was supported by the National Institutes of Health R01-MD012765, R01-DK117445, and R21-HL140385 to NF. The TRANSLATE studies and MT are supported by British Heart Foundation project grants PG/17/35/33001 and PG/19/16/34270, Kidney Research UK grant RP\_017\_20180302, and Medical University of Silesia grants KNW-1-152/N/7/K and KNW-1-171/N/6/K. WT206194. DLM and REM are supported by Alzheimer's Research UK major project grant ARUK-PG2017B-10. CH is supported by an MRC University Unit Programme Grant MC\_UU\_00007/10 (QTL in Health and Disease). SJL and MKL are supported by the Intramural Program of the NIH, National Institute of Environmental Health Sciences (ZO1 ES043012). SB acknowledges funding from the Wellcome Trust (218274/Z/19/Z). CEB and SIB are supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH. The CATHGEN study was supported by NIH grants HL095987 and HL036587. Generation Scotland (GS) received core support from the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HR03006). Genotyping and DNA methylation profiling of the GS samples was carried out by the Genetics Core Laboratory

at the Clinical Research Facility, University of Edinburgh, Edinburgh, Scotland, and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award “Stratifying Resilience and Depression Longitudinally” (STRADL; Reference 104036/Z/14/Z)).

Support for the Multi-Ethnic Study of Atherosclerosis (MESA) projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN2682015000031, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420. Also supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The TOPMed MESA Multi-Omics project was conducted by the University of Washington and LABioMed (HHSN2682015000031/HHSN26800004).

The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN2682018000131), Tougaloo College (HHSN2682018000141), the Mississippi State Department of Health (HHSN2682018000151), and the University of Mississippi Medical Center (HHSN2682018000101, HHSN2682018000111, and HHSN2682018000121) contracts from the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute for Minority Health and Health Disparities (NIMHD). The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute; the National Institutes of Health; or the US Department of Health and Human Services. The funders had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; and in the preparation, review, or approval of the manuscript.

The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, US Department of Health and Human Services, through contracts HHSN268201600018C, HHSN268201600001C, HHSN268201600002C, HHSN268201600003C, and HHSN268201600004C. A full listing of WHI investigators can be found at <http://www.whi.org/researchers/Documents%20Write%20a%20Paper/WHI%20Investigator%20List.pdf>.

The Hypertension Genetic Epidemiology Network (HyperGEN) Study is part of the National Heart, Lung, and Blood Institute (NHLBI) Family Blood Pressure Program; collection of the data represented here was supported by grants U01 HL054472, U01 HL054473, U01 HL054495, and U01 HL054509. The HyperGEN: Genetics of Left Ventricular Hypertrophy Study was supported by NHLBI grant R01 HL055673 with whole-genome sequencing made possible by supplement -18S1. The epigenetic data in HyperGEN was funded by an American Heart Association Cardiovascular Genome-Phenome Study Pathway Grant Award 15GSPSG23890000.

#### Availability of data and materials

Investigators interested in retrieving the controlled-access data at dbGap for WHI-BAA23 and MESA should apply using identifiers phs000200.v10.p3 and phs001416.v1.p, respectively (available online at [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs000200.v10.p3](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000200.v10.p3) and [https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs001416.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001416.v1.p1)). The summary results from this study are placed in dbGaP with accession number phs000930.v8.p1. While awaiting data release via dbGaP, data are available at [https://sph.unc.edu/wp-content/uploads/sites/112/2021/02/EWAS\\_COGENT.tar](https://sph.unc.edu/wp-content/uploads/sites/112/2021/02/EWAS_COGENT.tar). Access to the WHI-EMPC DNA methylation dataset is available upon request to [www.whi.org](http://www.whi.org). DNA methylation datasets from JHS and HyperGEN have been recently generated and upload to dbGap is ongoing. JHS data are available on request from [www.jacksonheartstudy.org](http://www.jacksonheartstudy.org) and HyperGEN data are available from the corresponding author from Ammous et al. [62]. Currently, JHS and WHI datasets are available through a scientific review application process directed to each respective study publication and presentation committee. Data can be obtained from the coordinating center of WHI and JHS after signing a data use agreement with the study. For research projects that meet the rules for access, CATHGEN data are available from the CATHGEN Steering Committee. Relevant requests are to be sent to the contact person at the CATHGEN Steering Committee, Melissa Hurdle ([melissa.hurdle@duke.edu](mailto:melissa.hurdle@duke.edu)). According to the terms of consent for GS participants,

access to individual-level data (omics and phenotypes) must be reviewed by the GS Access Committee. Applications should be made to [access@generationscotland.org](mailto:access@generationscotland.org). The source code used for the eFORGE analyses is publicly available at <https://github.com/charlesbreeze/eFORGE> [32] and <https://github.com/charlesbreeze/eFORGE-TF> [33].

#### Declarations

##### Ethics approval and consent to participate

All participants provided informed written consent to participate in these studies, and studies were approved by the Institutional Review Boards at each recruiting site.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department Health and Human Services, Bethesda, MD, USA. <sup>2</sup>UCL Cancer Institute, University College London, London WC1E 6BT, UK. <sup>3</sup>Altius Institute for Biomedical Sciences, Seattle, WA 98121, USA. <sup>4</sup>Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27516, USA. <sup>5</sup>Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, NC 27709, USA. <sup>6</sup>Division of Biomedical Informatics and Personalized Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, USA. <sup>7</sup>Division of Cardiovascular Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK. <sup>8</sup>Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, UK. <sup>9</sup>Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC 27701, USA. <sup>10</sup>Department of Biostatistics, University of Alabama, Birmingham, AL, USA. <sup>11</sup>Department of Public Health Sciences and Medicine, Loyola University Chicago, Maywood, IL, USA. <sup>12</sup>Division of Nephrology and Hypertension, Loyola University Chicago, Maywood, IL, USA. <sup>13</sup>Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. <sup>14</sup>Department of Pathology & Laboratory Medicine, Larner College of Medicine, University of Vermont, Burlington, VT, USA. <sup>15</sup>Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC, USA. <sup>16</sup>Department of Biochemistry, Larner College of Medicine, University of Vermont, Burlington, VT, USA. <sup>17</sup>Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA. <sup>18</sup>Division of Cardiology, Department of Medicine, School of Medicine, Duke University, Durham, NC, USA. <sup>19</sup>Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. <sup>20</sup>Center for Global Oncology, Institute of Global Health, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. <sup>21</sup>Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA. <sup>22</sup>Department of Medicine, University of North Carolina, Chapel Hill, NC 27599, USA. <sup>23</sup>School of Health and Life Sciences, Federation University Australia, Ballarat, VIC, Australia. <sup>24</sup>Department of Physiology, University of Melbourne, Parkville, VIC, Australia. <sup>25</sup>Department of Cardiovascular Sciences, University of Leicester, Leicester, UK. <sup>26</sup>Laboratory of Environmental Epigenetics, Departments of Environmental Health Sciences and Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA. <sup>27</sup>Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. <sup>28</sup>Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, The University of Manchester, Manchester, UK. <sup>29</sup>Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA. <sup>30</sup>College of Public Health, University of Kentucky, Lexington, KY, USA. <sup>31</sup>Durham VA Health System, Durham, NC 27705, USA. <sup>32</sup>The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA. <sup>33</sup>Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. <sup>34</sup>MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, UK. <sup>35</sup>Department of Human Genetics, David Geffen

School of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA. <sup>36</sup>Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, CA 90095, USA. <sup>37</sup>Manchester Heart Centre and Manchester Academic Health Science Centre, Manchester University NHS Foundation Trust, Manchester, UK.

Received: 10 July 2020 Accepted: 24 March 2021

Published online: 30 April 2021

## References

- GBD 2016 Causes of Death Collaborators. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017;390:1151–210.
- Saran R, Robinson B, Abbott KC, Bragg-Gresham J, Chen X, Gipson D, Gu H, Hirth RA, Hutton D, Jin Y, Kapke A, Kurtz V, Li Y, McCullough K, Modi Z, Morgenstern H, Mukhopadhyay P, Pearson J, Pisoni R, Repeck K, Schaubel DE, Shamraj R, Steffick D, Turf M, Woodside KJ, Xiang J, Yin M, Zhang X, Shahinian V. US renal data system 2019 annual data report: epidemiology of kidney disease in the United States. *Am J Kidney Dis*. 2020;75(1):A6–7. <https://doi.org/10.1053/j.ajkd.2019.09.003>.
- Saran R, Robinson B, Abbott KC, Agodoa LYC, Bhave N, Bragg-Gresham J, Balkrishnan R, Dietrich X, Eckard A, Eggers PW, Gaipov A, Gillen D, Gipson D, Hailpern SM, Hall YN, Han Y, He K, Herman W, Heung M, Hirth RA, Hutton D, Jacobsen SJ, Jin Y, Kalantar-Zadeh K, Kapke A, Kovesdy CP, Lavalley D, Leslie J, McCullough K, Modi Z, Molnar MZ, Montez-Rath M, Moradi H, Morgenstern H, Mukhopadhyay P, Nallamothu B, Nguyen DV, Norris KC, O'Hare AM, Obi Y, Park C, Pearson J, Pisoni R, Potukuchi PK, Rao P, Repeck K, Rhee CM, Schrager J, Schaubel DE, Selewski DT, Shaw SF, Shi JM, Shieu M, Sim JJ, Soohoo M, Steffick D, Streja E, Sumida K, Tamura MK, Tilea A, Tong L, Wang D, Wang M, Woodside KJ, Xin X, Yin M, You AS, Zhou H, Shahinian V. US renal data system 2017 annual data report: epidemiology of kidney disease in the United States. *Am J Kidney Dis*. 2018;71(3):A7. <https://doi.org/10.1053/j.ajkd.2018.01.002>.
- Morris AP, Le TH, Wu H, Akbarov A, van der Most PJ, Hemani G, et al. Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. *Nat Commun*. 2019;10(1):29. <https://doi.org/10.1038/s41467-018-07867-7>.
- Rakyan VK, Down TA, Balding DJ, Beck S. Epigenome-wide association studies for common human diseases. *Nat Rev Genet*. 2011;12(8):529–41. <https://doi.org/10.1038/nrg3000>.
- Taudt A, Colomé-Tatché M, Johannes F. Genetic sources of population epigenomic variation. *Nat Rev Genet*. 2016;17(6):319–32. <https://doi.org/10.1038/nrg.2016.45>.
- Qiu C, Hanson RL, Fufaa G, Kobes S, Gluck C, Huang J, Chen Y, Raj D, Nelson RG, Knowler WC, Susztak K. Cytosine methylation predicts renal function decline in American Indians. *Kidney Int*. 2018;93(6):1417–31. <https://doi.org/10.1016/j.kint.2018.01.036>.
- Smyth LJ, McKay GJ, Maxwell AP, McKnight AJ. DNA hypermethylation and DNA hypomethylation is present at different loci in chronic kidney disease. *Epigenetics*. 2014;9(3):366–76. <https://doi.org/10.4161/epi.27161>.
- Wing MR, Devaney JM, Joffe MM, Xie D, Feldman HI, Dominic EA, Guzman NJ, Ramezani A, Susztak K, Herman JG, Cope L, Harmon B, Kwabi-Addo B, Gordish-Dressman H, Go AS, He J, Lash JP, Kusek JW, Raj DS, for the Chronic Renal Insufficiency Cohort (CRIC) Study. DNA methylation profile associated with rapid decline in kidney function: findings from the CRIC study. *Nephrol Dial Transplant*. 2014;29(4):864–72. <https://doi.org/10.1093/ndt/gft537>.
- Richard MA, Huan T, Lighthart S, Gondalia R, Jhun MA, Brody JA, Irvin MR, Marioni R, Shen J, Tsai PC, Montasser ME, Jia Y, Syme C, Salfati EL, Boerwinkle E, Guan W, Mosley TH Jr, Bressler J, Morrison AC, Liu C, Mendelson MM, Uitterlinden AG, van Meurs JB, Franco OH, Zhang G, Li Y, Stewart JD, Bis JC, Psaty BM, Chen YDI, Kardia SLR, Zhao W, Turner ST, Absher D, Aslibekyan S, Starr JM, McRae AF, Hou L, Just AC, Schwartz JD, Vokonas PS, Menni C, Spector TD, Shuldiner A, Damcott CM, Rotter JJ, Palmas W, Liu Y, Paus T, Horvath S, O'Connell JR, Guo X, Pausova Z, Assimes TL, Sotoodehnia N, Smith JA, Arnett DK, Deary IJ, Baccarelli AA, Bell JT, Whitsel E, Dehghan A, Levy D, Fornage M, Heijmans BT, 't Hoen PAC, van Meurs J, Isaacs A, Jansen R, Franke L, Boomsma DI, Pool R, van Dongen J, Hottenga JJ, van Greevenbroek MMJ, Stehouwer CDA, van der Kallen CJH, Schalkwijk CG, Wijmenga C, Zernakova A, Tigchelaar EF, Slagboom PE, Beekman M, Deelen J, van Heemst D, Veldink JH, van den Berg LH, van Duijn CM, Hofman A, Uitterlinden AG, Jhamai PM, Verbiest M, Suchiman HED, Verkerk M, van der Breggen R, van Rooij J, Lakenberg N, Mei H, van Iterson M, van Galen M, Bot J, van 't Hof P, Deelen P, Nooren I, Moed M, Vermaat M, Zernakova DV, Luijk R, Bonder MJ, van Dijk F, Arindrarto W, Kielbasa SM, Swertz MA, van Zwet EW. DNA methylation analysis identifies loci for blood pressure regulation. *Am J Hum Genet*. 2017;101(6):888–902. <https://doi.org/10.1016/j.ajhg.2017.09.028>.
- Chu AY, Tin A, Schlosser P, Ko Y-A, Qiu C, Yao C, Joehanes R, Grams ME, Liang L, Gluck CA, Liu C, Coresh J, Hwang SJ, Levy D, Boerwinkle E, Pankow JS, Yang Q, Fornage M, Fox CS, Susztak K, Köttgen A. Epigenome-wide association studies identify DNA methylation associated with kidney function. *Nat Commun*. 2017;8(1):1286. <https://doi.org/10.1038/s41467-017-01297-7>.
- Sikdar S, Joehanes R, Joubert BR, Xu C-J, Vives-Usano M, Rezwan FI, Felix FJ, Ward JM, Guan W, Richmond RC, Brody JA, Küpers LK, Baiz N, Häberg SE, Smith JA, Reese SE, Aslibekyan S, Hoyo C, Dhingra R, Markunas CA, Xu T, Reynolds LM, Just AC, Mandaviya PR, Ghantous A, Bennett BD, Wang T, Consortium TBIO, Bakulski KM, Melen E, Zhao S, Jin J, Herceg Z, Meurs J, Taylor JA, Baccarelli AA, Murphy SK, Liu Y, Munthe-Kaas MC, Deary IJ, Nystad W, Waldenberger M, Annesi-Maesano I, Conneely K, Jaddoe VVW, Arnett D, Snieder H, Kardia SLR, Relton CL, Ong KK, Ewart S, Moreno-Macias GH, Romieu I, Sotoodehnia N, Fornage M, Moutsinger-Reif A, Koppelman GH, Bustamante M, Levy D, London SJ. Comparison of smoking-related DNA methylation between newborns from prenatal exposure and adults from personal smoking. *Epigenomics*. 2019;11:1487–1500, 13, doi: <https://doi.org/10.2217/epi-2019-0066>.
- Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative study group. *Control Clin Trials*. 1998; 19:61–109.
- Anderson GL, Manson J, Wallace R, Lund B, Hall D, Davis S, Shumaker S, Wang CY, Stein E, Prentice RL. Implementation of the Women's Health Initiative study design. *Ann Epidemiol*. 2003;13(9):S5–17. [https://doi.org/10.1016/S1047-2797\(03\)00043-7](https://doi.org/10.1016/S1047-2797(03)00043-7).
- Howard BV, Van Horn L, Hsia J, Manson JE, Stefanick ML, Wassertheil-Smolter S, et al. Low-fat dietary pattern and risk of cardiovascular disease: the Women's Health Initiative Randomized Controlled Dietary Modification Trial. *JAMA*. 2006;295(6):655–66. <https://doi.org/10.1001/jama.295.6.655>.
- Jackson RD, LaCroix AZ, Gass M, Wallace RB, Robbins J, Lewis CE, et al. Calcium plus vitamin D supplementation and the risk of fractures. *N Engl J Med*. 2006;354(7):669–83. <https://doi.org/10.1056/NEJMoa055218>.
- Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacob DR Jr, Kronmal R, Liu K, Nelson JC, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-ethnic study of atherosclerosis: objectives and design. *Am J Epidemiol*. 2002;156(9):871–81. <https://doi.org/10.1093/aje/kwf113>.
- Taylor HA, Wilson JG, Jones DW, Sarpong DF, Srinivasan A, Garrison RJ, et al. Toward resolution of cardiovascular health disparities in African Americans: design and methods of the Jackson Heart Study. *Ethn Dis*. 2005;15:S6–4–17.
- Akinremiju T, Do AN, Patki A, Aslibekyan S, Zhi D, Hidalgo B, Tiwari HK, Absher D, Geng X, Arnett DK, Irvin MR. Epigenome-wide association study of metabolic syndrome in African-American adults. *Clin Epigenetics*. 2018; 10(1):49. <https://doi.org/10.1186/s13148-018-0483-2>.
- Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, Deary IJ, MacIntyre DJ, Campbell H, McGilchrist M, Hocking LJ, Wisely L, Ford I, Lindsay RS, Morton R, Palmer CNA, Dominiczak AF, Porteous DJ, Morris AD. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol*. 2013;42(3):689–700. <https://doi.org/10.1093/ije/dys084>.
- Kraus WE, Granger CB, Sketch MH, Donahue MP, Ginsburg GS, Hauser ER, et al. A guide for a cardiovascular genomics biorepository: the CATHGEN experience. *J Cardiovasc Transl Res*. 2015;8(8):449–57. <https://doi.org/10.1007/s12265-015-9648-y>.
- Inker LA, Schmid CH, Tighiouart H, Eckfeldt JH, Feldman HI, Greene T, Kusek JW, Manzi J, van Lente F, Zhang YL, Coresh J, Levey AS, CKD-EPI Investigators. Estimating glomerular filtration rate from serum creatinine and cystatin C. *N Engl J Med*. 2012;367(1):20–9. <https://doi.org/10.1056/NEJMoa1114248>.
- Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, Beck S. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 2013;29(2):189–96. <https://doi.org/10.1093/bioinformatics/bts680>.
- Fortin J-P, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi.



- Bioinformatics. 2017;33(4):558–60. <https://doi.org/10.1093/bioinformatics/btw691>.
25. Touleimat N, Tost J. Complete pipeline for Infinium(®) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics*. 2012;4(3):325–41. <https://doi.org/10.2217/epi.12.21>.
  26. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics* 2013;14:293, 1, doi: <https://doi.org/10.1186/1471-2164-14-293>.
  27. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118–27. <https://doi.org/10.1093/biostatistics/kxj037>.
  28. Chen Y, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013;8(2):203–9. <https://doi.org/10.4161/epi.23470>.
  29. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13(1):86. <https://doi.org/10.1186/1471-2105-13-86>.
  30. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8):904–9. <https://doi.org/10.1038/ng1847>.
  31. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011;98(4):288–95. <https://doi.org/10.1016/j.ygeno.2011.07.007>.
  32. Breeze CE, Paul DS, van Dongen J, Butcher LM, Ambrose JC, Barrett JE, Lowe R, Rakyian VK, Iotchkova V, Frontini M, Downes K, Ouwehand WH, Laperle J, Jacques PÉ, Bourque G, Bergmann AK, Siebert R, Vellenga E, Saeed S, Matarese F, Martens JHA, Stunnenberg HG, Teschendorff AE, Herrero J, Birney E, Dunham I, Beck S. eFORGE: a tool for identifying cell type-specific signal in epigenomic data. *Cell Rep*. 2016;17(8):2137–50. <https://doi.org/10.1016/j.celrep.2016.10.059>.
  33. Breeze CE, Reynolds AP, van Dongen J, Dunham I, Lazar J, Neph S, et al. eFORGE v2.0: updated analysis of cell type-specific signal in epigenomic data. *Bioinformatics*. 2019;35:4767–9.
  34. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317–30.
  35. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc*. 2013; 8(8):1551–66. <https://doi.org/10.1038/nprot.2013.092>.
  36. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet*. 2017;49(1):131–8. <https://doi.org/10.1038/ng.3721>.
  37. Zhernakova DV, Deelen P, Vermaat M, van IJterson M, van Galen M, Arindartto W, van 't Hof P, Mei H, van Dijk F, Westra HJ, Bonder MJ, van Rooij J, Verkerk M, Jhamai PM, Moed M, Kielbasa SM, Bot J, Nooren I, Pool R, van Dongen J, Hottenga JJ, Stehouwer CDA, van der Kallen CJH, Schalkwijk CG, Zhernakova A, Li Y, Tigchelaar EF, de Klein N, Beekman M, Deelen J, van Heemst D, van den Berg LH, Hofman A, Uitterlinden AG, van Greevenbroek MMJ, Veldink JH, Boomsma DI, van Duijn CM, Wijmenga C, Slagboom PE, Swertz MA, Isaacs A, van Meurs JBJ, Jansen R, Heijmans BT, 't Hoen PAC, Franke L. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat Genet* 2017;49:139–145, 1, doi: <https://doi.org/10.1038/ng.3737>.
  38. Huan T, Joehanes R, Song C, Peng F, Guo Y, Mendelson M, Yao C, Liu C, Ma J, Richard M, Agha G, Guan W, Almlil LM, Conneely KN, Keefe J, Hwang SJ, Johnson AD, Fornage M, Liang L, Levy D. Genome-wide identification of DNA methylation QTLs in whole blood highlights pathways for cardiovascular disease. *Nat Commun*. 2019;10(1):4267. <https://doi.org/10.1038/s41467-019-12228-z>.
  39. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, Zheng J, Duggirala A, McArdle WL, Ho K, Ring SM, Evans DM, Davey Smith G, Relton CL. Systematic identification of genetic influences on methylation across the human life course. *Genome Biol*. 2016;17(1):61. <https://doi.org/10.1186/s13059-016-0926-z>.
  40. Xu X, Eales JM, Akbarov A, Guo H, Becker L, Talavera D, Ashraf F, Nawaz J, Pramanik S, Bowers J, Jiang X, Dormer J, Denniff M, Antczak A, Szulinska M, Wise I, Prestes PR, Glyda M, Bogdanski P, Zukowska-Szczechowska E, Berzuini C, Woolf AS, Samani NJ, Charchar FJ, Tomaszewski M. Molecular insights into genome-wide association studies of chronic kidney disease-defining traits. *Nat Commun*. 2018;9(1):4800. <https://doi.org/10.1038/s41467-018-07260-4>.
  41. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*. 2013;499(7456):43–9. <https://doi.org/10.1038/nature12222>.
  42. Rowland J, Akbarov A, Eales J, Xu X, Dormer JP, Guo H, Denniff M, Jiang X, Ranjzad P, Nazgiewicz A, Prestes PR, Antczak A, Szulinska M, Wise IA, Zukowska-Szczechowska E, Bogdanski P, Woolf AS, Samani NJ, Charchar FJ, Tomaszewski M. Uncovering genetic mechanisms of kidney aging through transcriptomics, genomics, and epigenomics. *Kidney Int*. 2019;95(3):624–35. <https://doi.org/10.1016/j.kint.2018.10.029>.
  43. Tomaszewski M, Eales J, Denniff M, Myers S, Chew GS, Nelson CP, Christofidou P, Desai A, Büsst C, Wojnar L, Musialik K, Jozwiak J, Debiec R, Dominiczak AF, Navis G, van Gilst WH, van der Harst P, Samani NJ, Harrap S, Bogdanski P, Zukowska-Szczechowska E, Charchar FJ. Renal mechanisms of association between fibroblast growth factor 1 and blood pressure. *J Am Soc Nephrol*. 2015;26(12):3151–60. <https://doi.org/10.1681/ASN.2014121211>.
  44. Hellwege JN, Velez Edwards DR, Giri A, Qiu C, Park J, Torstenson ES, Keaton JM, Wilson OD, Robinson-Cohen C, Chung CP, Roumie CL, Klarin D, Damrauer SM, DuVall SL, Siew E, Akwo EA, Wuttke M, Gorski M, Li M, Li Y, Gaziano JM, Wilson PWF, Tsao PS, O'Donnell CJ, Kovesdy CP, Pattaro C, Köttgen A, Susztak K, Edwards TL, Hung AM. Mapping eGFR loci to the renal transcriptome and phenotype in the VA million veteran program. *Nat Commun*. 2019;10(1):3842. <https://doi.org/10.1038/s41467-019-11704-w>.
  45. Ongen H, Buil A, Brown AA, Dermizakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics*. 2016; 32(10):1479–85. <https://doi.org/10.1093/bioinformatics/btv722>.
  46. Chebotareva N, Bobkova I, Shilov E. Heat shock proteins and kidney disease: perspectives of HSP therapy. *Cell Stress Chaperones*. 2017;22(3):319–43. <https://doi.org/10.1007/s12192-017-0790-0>.
  47. Wuttke M, Li Y, Li M, Sieber KB, Feitosa MF, Gorski M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet*. 2019;51(6):957–72. <https://doi.org/10.1038/s41588-019-0407-x>.
  48. Xu J, Liu H, Chai OH, Lan Y, Jiang R. Osr1 interacts synergistically with Wt1 to regulate kidney organogenesis. *PLoS One*. 2016;11(7):e0159597. <https://doi.org/10.1371/journal.pone.0159597>.
  49. Jiang H, Li L, Yang H, Bai Y, Jiang H, Li Y. Pax2 may play a role in kidney development by regulating the expression of TBX1. *Mol Biol Rep*. 2014; 41(11):7491–8. <https://doi.org/10.1007/s11033-014-3639-y>.
  50. Alam-Faruque Y, Hill DP, Dimmer EC, Harris MA, Foulger RE, Tweedie S, Attrill H, Howe DG, Thomas SR, Davidson D, Woolf AS, Blake JA, Mungall CJ, O'Donovan C, Apweiler R, Huntley RP. Representing kidney development using the gene ontology. *Plos One*. 2014;9(6):e99864. <https://doi.org/10.1371/journal.pone.0099864>.
  51. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, Klemm A, Flicek P, Manolio T, Hindorf L, Parkinson H. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014;42(D1): D1001–6. <https://doi.org/10.1093/nar/gkt1229>.
  52. Chen J, Huang Y, Hui Q, Mathur R, Gwinn M, So-Armah K, Freiberg MS, Justice AC, Xu K, Marconi VC, Sun YW. Epigenetic associations with estimated glomerular filtration rate among men with human immunodeficiency virus infection. *Clin Infect Dis*. 2020;70(4):667–73. <https://doi.org/10.1093/cid/ciz240>.
  53. Collins AJ, Foley RN, Gilbertson DT, Chen S-C. United States Renal Data System public health surveillance of chronic kidney disease and end-stage renal disease. *Kidney Int Suppl* (2011). 2015;5:2–7.
  54. Xue JL, Eggers PW, Agodoa LY, Foley RN, Collins AJ. Longitudinal study of racial and ethnic differences in developing end-stage renal disease among aged medicare beneficiaries. *J Am Soc Nephrol*. 2007;18(4):1299–306. <https://doi.org/10.1681/ASN.2006050524>.
  55. Collins AJ, Foley RN, Herzog C, Chavers B, Gilbertson D, Herzog C, et al. US Renal Data System 2012 annual data report. *Am J Kidney Dis*. 2013;61(A7): e1–476.
  56. Hoffmann TJ, Ehret GB, Nandakumar P, Ranatunga D, Schaefer C, Kwok P-Y, Iribarren C, Chakravarti A, Risch N. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat Genet*. 2017;49(1):54–64. <https://doi.org/10.1038/ng.3715>.

57. Imamura CK, Furihata K, Okamoto S, Tanigawara Y. Impact of cytochrome P450 2C19 polymorphisms on the pharmacokinetics of tacrolimus when coadministered with voriconazole. *J Clin Pharmacol*. 2016;56(4):408–13. <https://doi.org/10.1002/jcph.605>.
58. Ramírez V, Mejía-Vilet JM, Hernández D, Gamba G, Bobadilla NA. Radicol, a heat shock protein 90 inhibitor, reduces glomerular filtration rate. *Am J Physiol Renal Physiol*. 2008;295(4):F1044–51. <https://doi.org/10.1152/ajprenal.90278.2008>.
59. Lowe R, Slodkowitz G, Goldman N, Rakyán VK. The human blood DNA methylome displays a highly distinctive profile compared with other somatic tissues. *Epigenetics*. 2015;0, 10, 4, 274, 281, doi: <https://doi.org/10.1080/15592294.2014.1003744>.
60. Tsai P-C, Glastonbury CA, Eliot MN, Bollepalli S, Yet I, Castillo-Fernandez JE, Carnero-Montoro E, Hardiman T, Martin TC, Vickers A, Mangino M, Ward K, Pietiläinen KH, Deloukas P, Spector TD, Viñuela A, Loucks EB, Ollikainen M, Kelsey KT, Small KS, Bell JT. Smoking induces coordinated DNA methylation and gene expression changes in adipose tissue with consequences for metabolic health. *Clin Epigenetics*. 2018;10(1):126. <https://doi.org/10.1186/s13148-018-0558-0>.
61. Barker DJP, Bagby SP, Hanson MA. Mechanisms of disease: in utero programming in the pathogenesis of hypertension. *Nat Clin Pract Nephrol*. 2006;2(12):700–7. <https://doi.org/10.1038/ncpneph0344>.
62. Ammous F, Zhao W, Ratliff SM, Kho M, Shang L, Jones AC, Chaudhary NS, Tiwari HK, Irvin MR, Arnett DK, Mosley TH, Bielak LF, Kardia SLR, Zhou X, Smith J. Epigenome-wide association study identifies DNA methylation sites associated with target organ damage in older African Americans. *Epigenetics*. 2020;1–14. <https://doi.org/10.1080/15592294.2020.1827717>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

