OPEN ACCESS

University of Dundee

University of Dundee

A catalog of genetic loci associated with kidney function from analyses of a million individuals

LifeLines Cohort Study; Million Veteran Program; Wuttke, Matthias; Li, Yong; Li, Man; Sieber, Karsten B.

Published in:
Nature Genetics

DOI:
10.1038/s41588-019-0407-x

Publication date:
2019

Document Version
Peer reviewed version

Link to publication in Discovery Research Portal

Citation for published version (APA):
LifeLines Cohort Study, Million Veteran Program, Wuttke, M., Li, Y., Li, M., Sieber, K. B., Feitosa, M. F., Gorski, M., Tin, A., Wang, L., Chu, A. Y., Hoppmann, A., Kirsten, H., Giri, A., Chai, J-F., Sveinbjornsson, G., Tayo, B. O., Nutile, T., Fuchsberger, C., ... Pattaro, C. (2019). A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nature Genetics*, *51*(6), 957-972. https://doi.org/10.1038/s41588-019-0407-x

General rights
Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 28. Aug. 2022

A catalogue of molecular targets for kidney function from genetic analyses of a million individuals

Matthias Wuttke*[1], Yong Li*[1], Man Li*[2], Karsten B. Sieber*[3], Mary F. Feitosa*[4], Mathias Gorski*[5,6], Adrienne Tin[7,8], Lihua Wang[4], Audrey Chu[9], Anselm Hoppmann[1], Holger Kirsten[10,11], Ayush Giri[12,13], Jin-Fang Chai[14], Gardar Sveinbjornsson[15], Bamidele O. Tayo[16], Teresa Nutile[17], Christian Fuchsberger[18], Jonathan Marten[19], Massimiliano Cocca[20], Sahar Ghasemi[21,22], Yizhe Xu[2], Katrin Horn[10,11], Damia Noce[18], Peter J. van der Most[23], Sanaz Sedaghat[24], Zhi Yu[7,25], Masato Akiyama[26,27], Saima Afaq[28], Tarunveer S. Ahluwalia[29], Peter Almgren[30], Najaf Amin[24], Johan Ärnlöv[31,32], Stephan J.L. Bakker[33], Nisha Bansal[34,35], Daniela Baptista[36], Sven Bergmann[37,38,39], Mary L. Biggs[40,41], Ginevra Biino[42], Michael Boehnke[43,44], Eric Boerwinkle[45], Mathilde Boissel[46], Erwin P. Bottinger[47,48], Thibaud S. Boutin[19], Hermann Brenner[49,50], Marco Brumat[51], Ralph Burkhardt[11,52,53], Adam S. Butterworth[54,55], Eric Campana[56], Archie Campbell[57], Harry Campbell[58], Mickaël Canouil[46], Robert J. Carroll[59], Eulalia Catamo[20], John C. Chambers[28,60,61,62], Miao Ling Chee[63], Miao Li Chee[63], Xu Chen[64], Ching-Yu Cheng[63,65,66], Yurong Cheng[1], Kaare Christensen[67], Renata Cifkova[68,69], Marina Ciullo[17,70], Maria Pina Concas[20], James P. Cook[71], Josef Coresh[7], Tanguy Corre[38,72,73], Daniele Cusi[74,75], Sala Cinzia Felicita[76], John Danesh[77], E. Warwick Daw[4], Martin H. de Borst[78], Alessandro De Grandi[18], Renée de Mutsert[79], Aiko P.J. de Vries[80], Frauke Degenhardt[81], Graciela Delgado[82], Ayse Demirkan[24], Emanuele di Angelantonio[83,84,85], Katalin Dittrich[86,87], Jasmin Divers[88], Rajkumar Dorajoo[89], Kai-Uwe Eckardt[90], Georg Ehret[91], Paul Elliott[92,93,94,95], Karlhans Endlich[96,97], Michele K. Evans[98], Janine F. Felix[24,99,100], Valencia Foo[63], Oscar H. Franco[24,101], Andre Franke[81], Barry I. Freedman[102], Sandra Freitag-Wolf[103], Yechiel Friedlander[104], Philippe Froguel[46,105], Ron T. Gansevoort[106], He Gao[92], Paolo Gasparini[20,107], John Michael Gaziano[108], Vilmantas Giedraitis[109], Christian Gieger[110], Giorgia Girotto[20], Franco Giulianini[111], Martin Gögele[18], Scott D. Gordon[112], Vilmundur Gudnason[113,114], Toomas Haller[115], Pavel Hamet[116,117], Tamara B. Harris[118], Catharina A. Hartman[119], Thomsen Hauke[120], Caroline Hayward[19], Jacklyn N. Hellwege[121,122], Chew-Kiat Heng[123,124,125], Andrew A. Hicks[18], Edith Hofer[126,127], Wei Huang[128,129], Nina Hutri-Kähönen[130,131], Shih-Jen Hwang[132,133], M. Arfan Ikram[24], Olafur S. Indridason[134], Erik Ingelsson[135,136,137], Marcus Ising[138], Vincent W.V. Jaddoe[24,99,100], Johanna Jakobsdottir[113], Jost B. Jonas[139], Peter K. Joshi[58], Navya Shilpa Josyula[140], Bettina Jung[5], Mika Kähönen[141,142], Yoichiro Kamatani[26,143], Candace M. Kammerer[144], Masahiro Kanai[26,145], Mika Kastarinen[146], Shona M. Kerr[19], Chiea Chuen Khor[89,147], Wieland Kiess[11,86,87], Marcus E. Kleber[82], Wolfgang Koenig[148,149], Jaspal S. Kooner[62], Antje Körner[11,86,150], Peter Kovacs[151], Aldi T. Kraja[4], Alena Krajcoviechova[68,69], Holly Kramer[16,152], Bernhard K. Krämer[82], Florian Kronenberg[153], Michiaki Kubo[154], Brigitte Kühnel[110], Mikko Kuokkanen[155,156], Johanna Kuusisto[157,158], Markku Laakso[157,158], Martina La Bianca[20], Leslie A. Lange[159], Carl D. Langefeld[88], Jeannette Lee[14], Benjamin Lehne[28], Terho Lehtimäki[160,161], Wolfgang Lieb[162], Su-Chi Lim[163,164], Lars Lind[165], Cecilia M. Lindgren[166,167], Jun Liu[24], Jianjun Liu[89,168], Markus Loeffler[10,11], Ruth J.F. Loos[47,169], Susanne Lucae[138], Mary Ann Lukas[170], Leo-Pekka Lyytikäinen[160,161], Reedik Mägi[115], Patrik K.E. Magnusson[64], Anubha Mahajan[171,172], Nicholas G. Martin[112], Jade Martins[173], Winfried März[174,175,176], Deborah Mascalzoni[18], Koichi Matsuda[177], Christa Meisinger[178,179,180,181], Thomas Meitinger[182,183], Olle Melander[184], Andres Metspalu[115], Evgenia K. Mikaelsdottir[15], Yuri Milaneschi[185], Kozeta Miliku[24,99,100], Pashupati P. Mishra[142,160], Karen L. Mohlke[186], Nina Mononen[142,160], Grant W. Montgomery[187], Dennis O. Mook-Kanamori[188], Josyf C. Mychaleckyj[189], Girish N. Nadkarni[47,190], Mike A. Nalls[191,192], Matthias Nauck[22,193], Kjell Nikus[194,195], Boting Ning[196], Ilja M. Nolte[23], Raymond Noordam[197], Jeffrey O'Connell[198], Michelle L. O'Donoghue[199,200], Isleifur Olafsson[201], Albertine J. Oldehinkel[33], Marju Orho-Melander[30], Willem H. Ouwehand[77], Sandosh Padmanabhan[202], Nicholette D. Palmer[203],

Runolfur Palsson[114,134], Brenda W.J.H. Penninx[185], Thomas Perls[204], Markus Perola[205,206], Mario Pirastu[207], Nicola Pirastu[58], Giorgio Pistis[208], Anna I. Podgornaia[209], Ozren Polasek[210,211], Belen Ponte[212], David J. Porteous[57,213], Tanja Poulain[11], Peter P. Pramstaller[18], Michael H. Preuss[47], Bram P. Prins[54], Michael A. Province[4], Ton J. Rabelink[214], Laura M. Raffield[186], Olli T. Raitakari[215,216], Dermot F. Reilly[209], Rainer Rettig[217], Myriam Rheinberger[5], Kenneth M. Rice[41], Paul M. Ridker[111,218], Fernando Rivadeneira[24,219], Federica Rizzi[220,221], David J. Roberts[222], Antonietta Robino[20], Peter Rossing[29], Igor Rudan[58], Rico Rueedi[38,223], Daniela Ruggiero[17,70], Kathleen A. Ryan[224], Yasaman Saba[225], Charumathi Sabanayagam[63,65], Veikko Salomaa[205], Erika Salvi[220,226], Kai-Uwe Saum[49], Helena Schmidt[227], Reinhold Schmidt[126], Ben Schöttker[49,50], Christina-Alexandra Schulz[30], Nicole Schupf[228,229,230], Christian M. Shaffer[59], Yuan Shi[63,231], Albert V. Smith[232], Blair H. Smith[233], Nicole Soranzo[234], Cassandra N. Spracklen[186], Kari Stefansson[15], Konstantin Strauch[235,236], Heather M. Stringham[43,44], Michael Stumvoll[237], Per O. Svensson[238,239], Silke Szymczak[103], E-Shyong Tai[164,240,241], Salman M. Tajuddin[98], Nicholas Y. Q. Tan[63], Kent Taylor[242], Andrej Teren[11,243], Yih-Chung Tham[63], Joachim Thiery[11,52], Chris H.L. Thio[23], Gudmar Thorleifsson[15], Daniela Toniolo[76], Anke Tönjes[237], Johanne Tremblay[116,244], Ioanna Tzoulaki[92,245], André G. Uitterlinden[219], Simona Vaccargiu[207], Rob M. van Dam[14,168,246], Pim van der Harst[247,248,249], Cornelia M. van Duijn[24], Digna Velez Edward[12,13], Niek Verweij[247], Suzanne Vogelezang[24,99,100], Uwe Völker[97,250], Peter Vollenweider[251], Gerard Waeber[251], Melanie Waldenberger[110,178,252], Lars Wallentin[253,254], Ya Xing Wang[255], Chaolong Wang[89,256], Dawn M. Waterworth[3], Wen Bin Wei[257], Harvey White[258], John B. Whitfield[259], Sarah H. Wild[260], James F. Wilson[58], Mary K. Wojczynski[4], Charlene Wong[66], Tien Yin Wong[63,241,261], Liang Xu[255], Qiong Yang[196], Masayuki Yasuda[63,262], Laura M. Yerges-Armstrong[3], Weihua Zhang[28,263], Alan B. Zonderman[98], Jerome I. Rotter[264,265,266], Murielle Bochud[72], Bruce M. Psaty[40,267], Veronique Vitart[19], James G. Wilson[268], Abbas Dehghan[28,92], Afshin Parsa[269,270], Daniel I. Chasman[111,218], Kevin Ho[271,272], Andrew P. Morris[71,273], Olivier Devuyst[274], Shreeram Akilesh[275,276], Sarah A. Pendergrass[277], Xueling Sim[14], Carsten A. Böger[5,278], Yukinori Okada[279,280], Todd L. Edwards[281], Harold Snieder[23], Daniel F. Gudbjartsson[15], Adriana Hung[282,283], Iris M. Heid**[6], Markus Scholz**[10,11], Alexander Teumer**[21,22], Anna Köttgen**† [1,7], Cristian Pattaro**† [18]

* Indicates joint contribution

** Indicates joint oversight

† Indicates corresponding author

Authors for Correspondence:

Anna Köttgen, MD MPH
Institute of Genetic Epidemiology
Medical Center - University of Freiburg
Hugstetter Str. 49, 79106 Freiburg, Germany
Tel: +49 (0)761 270-78050
anna.koettgen@uniklinik-freiburg.de

Cristian Pattaro PhD
Eurac Research
Institute for Biomedicine
Via Galvani 31, 39100 Bolzano, Italy
+39 0471 055 527
cristian.pattaro@eurac.edu

Author affiliations

1 Institute of Genetic Epidemiology, Department of Biometry, Epidemiology and Medical Bioinformatics, Faculty of Medicine and Medical Centre - University of Freiburg, Freiburg, Germany
2 Department of Medicine, Division of Nephrology and Hypertension, University of Utah, Salt Lake City, USA
3 Target Sciences - Genetics, GlaxoSmithKline, Collegeville (Pennsylvania), USA
4 Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis (Missouri), USA
5 Department of Nephrology, University Hospital Regensburg, Regensburg, Germany
6 Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany
7 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore (Maryland), USA
8 Epidemiology and Clinical Research, Welch Centre for Prevention, Baltimore (Maryland), USA
9 Merck Research Laboratory, Merck & Co., Inc., USA
10 Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany
11 LIFE Research Centre for Civilization Diseases, University of Leipzig, Leipzig, Germany
12 Department of Obstetrics and Gynecology, Institute for Medicine and Public Health, Vanderbilt University Medical Centre, Nashville, USA
13 Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, USA
14 Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore
15 deCODE Genetics, Amgen Inc., Reykjavik, Iceland
16 Department of Public Health Sciences, Loyola University Chicago, Maywood (Illinois), USA
17 Institute of Genetics and Biophysics "Adriano Buzzati-Traverso" - CNR, Naples, Italy
18 Eurac Research, Institute for Biomedicine (affiliated to the University of Lübeck), Bolzano, Italy
19 Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
20 Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, Italy
21 Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
22 DZHK (German Centre for Cardiovascular Research), Germany
23 Department of Epidemiology, University Medical Centre Groningen, Groningen, Netherlands
24 Department of Epidemiology, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, Netherlands
25 Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore (Maryland), USA
26 Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan
27 Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
28 Department of Epidemiology and Biostatistics, Faculty of Medicine, School of Public Health, Imperial College London, London, UK
29 Steno Diabetes Centre Copenhagen, Gentofte, Denmark
30 Diabetes and Cardiovascular Disease - Genetic Epidemiology, Department of Clincial Sciences in Malmö, Lund University, Malmö, Sweden
31 Department of Neurobiology, Care Sciences and Society, Division of Family Medicine and Primary Care, Karolinska Institutet, Stockholm, Sweden
32 School of Health and Social Studies, Dalarna University, Sweden
33 Department of Internal Medicine, University Medical Centre Groningen, Groningen, Netherlands
34 Division of Nephrology, University of Washington, Seattle (Washington), USA
35 Kidney Research Institute, University of Washington, Seattle (Washington), USA
36 Cardiology, Geneva University Hospitals, Genève, Switzerland
37 University of Lausanne, Lausanne, Switzerland
38 Swiss Institute of Bioinformatics, Lausanne, Switzerland
39 Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa
40 Cardiovascular Health Research Unit, Department of Medicine, Department of Epidemiology, Department of Health Service, University of Washington, Seattle (Washington), USA
41 Department of Biostatistics, University of Washington, Seattle (Washington), USA
42 Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy
43 Department of Biostatistics, University of Michigan, Ann Arbor, USA
44 Centre for Statistical Genetics, University of Michigan, USA
45 Human Genetics Centre, University of Texas Health Science Centre, Houston (Texas), USA
46 CNRS UMR 8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, University of Lille, Lille, France
47 The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York (New York), USA
48 Digital Health Centre, Hasso Plattner Institute and University of Potsdam, Potsdam, Germany
49 Division of Clinical Epidemiology and Aging Research, German Cancer Research Centre, Heidelberg, Germany
50 Network Aging Research, University of Heidelberg, Heidelberg, Germany
51 Surgery and Health Sciences, Department of Medicine, University of Trieste, Trieste, Italy
52 Institute for Laboratory Medicine, University of Leipzig, Leipzig, Germany
53 Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg, Germany
54 MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
55 National Institute for Health Research Blood and Transplant Research Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK
56 Surgery and Health Sciences, Department of Medicine, University of Trieste, Mondovì, Italy
57 Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

58 Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
59 Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville (Tennessee), USA
60 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
61 Department of Cardiology, London North West Healthcare NHS Trust, Ealing Hospital, London, UK
62 MRC-PHE Centre for Environment and Health, Imperial College London, London, UK
63 Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore
64 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
65 Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore
66 Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
67 Unit of Epidemiology, Biostatistics and Biodemography, Department of Public Health, Southern Denmark University, Odense, Denmark
68 Centre for Cardiovascular Prevention, First Faculty of Medicine, Department of Medicine, Charles University in Prague, Prague, Czech Republic
69 Thomayer Hospital, Prague, Czech Republic
70 IRCCS Neuromed, Pozzilli, Italy
71 Department of Biostatistics, University of Liverpool, Liverpool, UK
72 Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland
73 Department of Computational Biology, University of Lausanne, Switzerland
74 Institute of Biomedical Technologies, Italy National Research Council, Bresso (Milano), Italy
75 Bio4Dreams - business nursery for life sciences, Bresso (Milano), Italy
76 San Raffaele Research Institute, Milano, Italy
77 University of Cambridge, Cambridge, UK
78 Department of Internal Medicine, Division of Nephrology, University Medical Centre Groningen, Groningen, Netherlands
79 Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, Netherlands
80 Section of Nephrology, Department of Internal Medicine, Leiden University Medical Centre, Leiden, Netherlands
81 Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany
82 Vth Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology, Diabetology), Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
83 Blood and Transplant Research Unit in Donor Health and Genomics, National Institute of Health Research, Cambridge, UK
84 Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
85 NHS Blood and Transplant, Cambridge, UK
86 Department of Women and Child Health, Hospital for Children and Adolescents, University of Leipzig, Leipzig, Germany
87 Centre for Pediatric Research, University of Leipzig, Leipzig, Germany
88 Public Health Sciences - Biostatistics, Wake Forest School of Medicine, Winston-Salem (North Carolina), USA
89 Genome Institute of Singapore, Agency for Science Technology and Research, Singapore, Singapore
90 Medical Department, Division of Nephrology and Internal Intensive Care Medicine CVK/CCM, Charité Universität Medizin Berlin, Germany
91 Cardiology, Geneva University Hospitals, Geneva, Switzerland
92 Department of Biostatistics and Epidemiology, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK
93 Imperial College NIHR Biomedical Research Centre, Imperial College London, London, UK
94 Dementia Research Institute, Imperial College London, London, UK
95 Health Data Research UK-London, London, UK
96 Department of Anatomy and Cell Biology, University Medicine Greifswald, Greifswald, Germany
97 DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, Germany
98 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Baltimore (Maryland), USA
99 The Generation R Study Group, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, Netherlands
100 Department of Pediatrics, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, Netherlands
101 Institute of Social and Preventive Medicine (ISPM), University of Bern, Bern, Switzerland
102 Internal Medicine - Section on Nephrology, Wake Forest School of Medicine, Winston-Salem (North Carolina), USA
103 Institute of Medical Informatics and Statistics, Kiel University, Kiel, Germany
104 School of Public Health and Community Medicine, Hebrew University of Jerusalem, Jerusalem, Israel
105 Department of Genomics of Common Disease, Imperial College London, London, UK
106 University Medical Centre Groningen, Groningen, Netherlands
107 DSM, University of Trieste, Trieste, Italy
108 Massachusetts Area Veterans Epidemiology Research and Information Centre (MAVERIC), VA Cooperative Studies Program, VA Boston Healthcare System, Boston (Massachusetts), USA
109 Department of Public Health and Caring Sciences, Molecular Geriatrics, Uppsala University, Uppsala, Sweden
110 Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany
111 Division of Preventive Medicine, Brigham and Women's Hospital, Boston, USA
112 QIMR Berghofer Medical Research Institute, Brisbane, Australia
113 Icelandic Heart Association, Kopavogur, Iceland
114 Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland
115 Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
116 Montreal University Hospital Research Centre, CHUM, Montreal, Canada
117 Medpharmgene, Montreal, Canada

118 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Bethesda (Maryland), USA
119 Interdisciplinary Centre Psychopathology and Emotion regulation (ICPE), University Medical Centre Groningen, Groningen, Netherlands
120 Division of Molecular Genetic Epidemiology, German Cancer Research Centre, Heidelberg, Germany
121 Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, Nashville, USA
122 Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, USA
123 Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
124 National University Children's Medical Institute, Khoo Teck Puat, Singapore
125 National University Health System, Singapore, Singapore
126 Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria
127 Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria
128 Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Centre, Shanghai, China
129 Shanghai Industrial Technology Institute, Shanghai, China
130 Department of Pediatrics, Tampere University Hospital, Tampere, Finland
131 Department of Pediatrics, Faculty of Medicine and Life Sciences, University of Tampere, Finland
132 NHLBI's Framingham Heart Study, Framingham (Massachusetts), USA
133 The Centre for Population Studies, NHLBI, Framingham (Massachusetts), USA
134 Division of Nephrology, Internal Medicine Services, Landspitaliâ€“The National University Hospital of Iceland, Reykjavik, Iceland
135 Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, USA
136 Stanford Cardiovascular Institute, Stanford University, USA
137 Molecular Epidemiology and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden
138 Max Planck Institute of Psychiatry, Munich, Germany
139 Department of Ophthalmology, Medical Faculty Mannheim, University Heidelberg, Mannheim, Germany
140 Geisinger Research, Biomedical and Translational Informatics Institute, Rockville, USA
141 Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland
142 Department of Clinical Chemistry, Finnish Cardiovascular Research Centre, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland
143 Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan
144 Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh (Pennsylvania), USA
145 Department of Biomedical Informatics, Harvard Medical School, Boston, USA
146 Kuopio University Hospital, Kuopio, Finland
147 Singapore Eye Research Institute, Singapore National Eye Centre, Singapore
148 Deutsches Herzzentrum München, Technische Universität München, Munich, Germany
149 Abteilung Innere Medzin II -Kardiologie, Universitätsklinikum Ulm, Ulm, Germany
150 Centre for Pediatric Research, University of Leipzig, Germany
151 Integrated Research and Treatment Centre Adiposity Diseases, University of Leipzig, Leipzig, Germany
152 Division of Nephrology and Hypertension, Loyola University Chicago, USA
153 Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria
154 RIKEN Centre for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan
155 The Department of Public Health Solutions, National Institute for Health and Welfare, Helsinki, Finland
156 Diabetes and Obesity Research Program, University of Helsinki, Helsinki, Finland
157 University of Eastern Finland, Kuopio, Finland
158 Kuopio University Hospital, Finland
159 Division of Biomedical Informatics and Personalized Medicine, School of Medicine, University of Colorado Denver - Anschutz Medical Campus, Aurora (Colorado), USA
160 Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland
161 Department of Clinical Chemistry, Finnish Cardiovascular Research Centre, Faculty of Medicine and Life Sciences, University of Tampere, Finland
162 Institute of Epidemiology and Biobank Popgen, Kiel University, Kiel, Germany
163 Diabetes Centre, Khoo Teck Puat Hospital, Singapore, Singapore
164 Saw Swee Hock School of Public Health, National University of Singapore, Singapore
165 Cardiovascular Epidemiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden
166 Nuffield Department of Medicine, University of Oxford, Oxford, UK
167 Broad Institute of Harvard and MIT, USA
168 Yong Loo Lin School of Medicine, National University of Singapore, Singapore
169 The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York (New York), USA
170 Target Sciences - Genetics, GlaxoSmithKline, Albuquerque (New Mexico), USA
171 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
172 Endocrinology and Metabolism, Oxford Centre for Diabetes, University of Oxford, UK
173 Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany
174 Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany
175 Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Austria
176 Medical Clinic V, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
177 Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan

178 Institute of Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany
179 Independent Clinical Epidemiology Research Group, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany
180 Ludwig-Maximilians-Universität München, Munich, Germany
181 Epidemiology, UNIKA-T Augsburg, Augsburg, Germany
182 Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany
183 Institute of Human Genetics, Technische Universität München, Munich, Germany
184 Hypertension and Cardiovascular Disease , Department of Clincial Sciences Malmö, Lund University, Malmö, Sweden
185 Department of Psychiatry, VU University Medical Centre, Amsterdam, Netherlands
186 Department of Genetics, University of North Carolina, Chapel Hill (North Carolina), USA
187 University of Queensland, St Lucia, Australia
188 Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, Netherlands
189 Centre for Public Health Genomics, University of Virginia, Charlottesville (Virginia), USA
190 Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York (New York), USA
191 Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda (Maryland), USA
192 Data Tecnica International, Glen Echo (Maryland), USA
193 Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany
194 Department of Cardiology, Heart Centre, Tampere University Hospital, Tampere, Finland
195 Department of Cardiology, Finnish Cardiovascular Research Centre - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland
196 Department of Biostatistics, Boston University School of Public Health, Boston (Massachusetts), USA
197 Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Centre, Leiden, Netherlands
198 University of Maryland, College Park, USA
199 Cardiovascular Division, Brigham and Women's Hospital, Boston, USA
200 TIMI Study Group, USA
201 Department of Clinical Biochemistry, Landspitali University Hospital, Reykjavik, Iceland
202 Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK
203 Biochemistry, Wake Forest School of Medicine, Winston-Salem (North Carolina), USA
204 Department of Medicine, Geriatrics Section, Boston Medical Center, Boston University School of Medicine, Boston (Massachusetts), USA
205 National Institute for Health and Welfare, Helsinki, Finland
206 The Diabetes and Obesity Research Program, University of Helsinki, Helsinki, Finland
207 Institute of Genetic and Biomedical Research, National Research Council of Italy, UOS of Sassari, Li Punti (Sassari), Italy
208 Department of Psychiatry, University Hospital of Lausanne, Lausanne, Switzerland
209 Merck & Co., Inc., Kenilworth (New Jersey), USA
210 Faculty of Medicine, University of Split, Split, Croatia
211 Gen-info Ltd, Zagreb, Croatia
212 Service de Néphrologie, Geneva University Hospitals, Geneva, Switzerland
213 Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK
214 Einthoven Laboratory of Experimental Vascular Research, Leiden University Medical Centre, Leiden, Netherlands
215 Turku University Hospital, Turku, Finland
216 University of Turku, Finland
217 Institute of Physiology, University Medicine Greifswald, Karlsburg, Germany
218 Harvard Medical School, USA
219 Department of Internal Medicine, Erasmus MC, University Medical Centre Rotterdam, Rotterdam, Netherlands
220 Department of Health Sciences, University of Milan, Milano, Italy
221 ePhood Scientific Unit, ePhood SRL, Milano, Italy
222 Nuffield Division of Clinical Laboratory Sciences, University of Oxford, Oxford, UK
223 Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
224 Division of Endocrinology Diabetes and Nutrition, School of Medicine, University of Maryland, College Park, USA
225 Molecular Biology and Biochemistry, Gottfried Schatz Research Centre for Cell Signaling, Metabolism and Aging, Medical University of Graz, Graz, Austria
226 Neuroalgology Unit, IRCCS Foundation â€œCarlo Bestaâ€□ Neurological Institute, Milano, Italy
227 Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University of Graz, Graz, Austria
228 Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, USA
229 Gertrude H. Sergievsky Centre, Columbia University Medical Centre, USA
230 Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University Medical Centre, New York, USA
231 Duke-NUS Medical School, Singapore
232 Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavík, Iceland
233 Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK
234 Wellcome Sanger Institute, South Cambridgeshire, UK
235 Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany
236 Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Munich, Germany
237 Department of Endocrinology and Nephrology, University of Leipzig, Leipzig, Germany
238 Department of Clinical Science and Education, Karolinska Institutet, Södersjukhuset (KI SÖS), Stockholm, Sweden
239 Department of Cardiology, Södersjukhuset, Sweden
240 Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
241 Duke-NUS Medical School, Singapore, Singapore

1 242 Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research institute, Harbor-UCLA
2 Medical Centre, Torrance (California), USA
3 243 Heart Centre Leipzig, Leipzig, Germany
4 244 CRCHUM, Montreal, Canada
5 245 Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece
6 246 Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, USA
7 247 Department of Cardiology, University Medical Centre Groningen, Groningen, Netherlands
8 248 Department of Genetics, University Medical Centre Groningen, Netherlands
9 249 Durrer Centre for Cardiovascular Research, Netherlands Heart Institute, Utrecht, Netherlands
10 250 Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany
11 251 Internal Medicine, Department of Medicine, Lausanne University Hospital, Lausanne, Switzerland
12 252 DZHK (German Centre for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich, Germany
13 253 Cardiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden
14 254 Uppsala Clinical Research Centre, Sweden
15 255 Beijing Institute of Ophthalmology, Beijing Key Laboratory of Ophthalmology and Visual Sciences, Beijing Tongren Hospital,
16 Capital Medical University, Beijing, China
17 256 School of Public Health, Tongji Medical School, Huazhong University of Science and Technology, China
18 257 Beijing Tongren Eye Centre, Beijing Tongren Hospital, Capital Medical University, Beijing, China
19 258 Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland, New Zealand
20 259 QIMR Berghofer Medical Research Institute, Queensland, Australia
21 260 Centre for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh,
22 Edinburgh, UK
23 261 Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
24 262 Department of Ophthalmology, Tohoku University Graduate School of Medicine, Japan
25 263 Department of Cardiology, Ealing Hospital, UK
26 264 Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research institute, Harbor-UCLA
27 Medical Centre, Torrance, USA
28 265 Department of Pediatrics, Harbor-UCLA Medical Centre, USA
29 266 Department of Medicine, Harbor-UCLA Medical Centre, Torrance, USA
30 267 Kaiser Permanente Washington Health Research Institute, Seattle (Washington), USA
31 268 Department of Physiology and Biophysics, University of Mississippi Medical Centre, Jackson (Mississippi), USA
32 269 Division of Kidney, Urologic and Hematologic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases,
33 National Institutes of Health, Bethesda, USA
34 270 Department of Medicine, University of Maryland, College Park, USA
35 271 Kidney Health Research Institute (KHRI), Geisinger, Danville (Pennsylvania), USA
36 272 Department of Nephrology, Geisinger, Danville (Pennsylvania), USA
37 273 Wellcome Trust Centre for Human Genetics, University of Oxford, UK
38 274 Institute of Physiology, University of Zurich, Zurich, Switzerland
39 275 Renal Pathology, University of Washington Medical Centre, Seattle, USA
40 276 Kidney Health Research Institute (KHRI), Seattle, USA
41 277 Geisinger Research, Biomedical and Translational Informatics Institute, Danville (Pennsylvania), USA
42 278 Department of Nephrology and Rheumatology, Kliniken Südostbayern AG, Regensburg, Germany
43 279 Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences (IMS), Osaka, Japan
44 280 Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan
45 281 Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, Nashville
46 (TN), USA
47 282 Vanderbilt University Medical Centre, Nashville (TN), USA
48 283 Department of Veterans Affairs, USA

1 **Abstract**

2 Chronic kidney disease (CKD) is a worldwide public health concern with multi-systemic
3 complications. We analyzed kidney function biomarkers combining 121 genome-wide
4 association studies across 5 ancestries and >1 million individuals in total. We identified 308 loci
5 that explained 20% of glomerular filtration rate's heritability. Fine-mapping analysis highlighted
6 missense driver variants in *CACNA1S*, *CPS1*, *CERS2*, *C9*, *EDEM3*, *KLHDC7A*, *PPM1J*,
7 *RPL3L*, *SLC22A2*, *SLC25A45*, and *SLC47A1*. Gene expression co-localization analyses in 46
8 human tissues including kidney compartments identified 20 target genes in the kidney, including
9 *UMOD*, *KNG1*, *MLLT3*, and *GALNTL5*. These results provide a catalogue of molecular targets
10 for translational research on CKD.

Chronic kidney disease (CKD) has a high worldwide prevalence.[1] It has limited therapeutic options, ranks among the leading causes of death,[2] and is associated with high costs for health systems.[3] Despite persons with CKD being at increased risk of multisystem complications, end-stage renal disease, cardiovascular disease, and early mortality, public and clinical awareness remain low.[4] Clinical trials in nephrology are still underrepresented compared to other disciplines,[5] and, with the exception of *SGLT2* inhibitors to slow progression of diabetic kidney disease,[6] no new effective treatments have become available over the last decade. A better understanding of the mechanisms underlying kidney function in health and disease can help identify much needed targets for drug development.

Genome-wide association studies (GWAS) and exome-chip studies of the glomerular filtration rate estimated from serum creatinine (eGFR), the main biomarker to quantify kidney function and define CKD, have shed light on the underlying mechanisms of CKD.[7] Nearly one hundred genetic loci were identified in samples of European[8-14], Asian[15-17], and across[18] ancestries. Similar to other complex traits and diseases, identifying causal genes and molecular mechanisms implicated by genetic associations represents a substantial challenge and has only been successful for few kidney function-associated loci.[19,20] Advanced statistical fine-mapping approaches and newly emerging gene expression data across a wide range of tissues open up new opportunities for prioritizing putative causal variants, effector genes, and target tissues, based on the results from large-scale GWAS meta-analyses.

In addition, the detection of co-expressed genes as well as gene sets, cell-type specific regulatory marks and pathways that are enriched for trait-associated signals is now possible, but requires particularly large GWAS sample sizes. The largest published GWAS meta-analyses of eGFR included up to 140,000 individuals.[8,9,21] However, the identified index variants only explained <4% of the eGFR variance.[8,9] A substantial expansion of study sample size, inclusion of more diverse populations, and more comprehensive coverage of genetic variants promise to identify novel loci, increase the explained variance of eGFR, and detect disease-relevant pathways and co-regulation with other complex traits.

We therefore carried out a trans-ethnic meta-analysis of GWAS of eGFR from 765,348 individuals. Generalizability of results was evaluated in an independent study of 280,722 individuals, for a combined sample size of >1 million participants. Associated loci were characterized through use of a complementary kidney function marker, blood urea nitrogen (BUN). To identify most likely causal variants, genes and mechanisms, we performed enrichment and network analyses, statistical fine-mapping, and integration of gene expression

in kidney and 44 other tissues. The resulting list of functionally relevant variants, genes, tissues and pathways provides a rich resource of potential therapeutic targets to improve CKD treatment and prevention.

Results

Overview

Within the CKD Genetics (CKDGen) Consortium, we established a collaborative, standardized and automated analysis workflow to integrate results from 121 GWAS of five ancestry groups (**Supplementary Table 1**). Our effort served two purposes (**Supplementary Figure 1**): first, we aimed at characterizing the genetics of kidney function at unprecedented breadth through meta-analysis of trans-ethnic samples; second, we aimed to understand each locus in depth among European ancestry (EA) individuals, for whom large reference panels on linkage disequilibrium (LD) structure are available.

Identification of 308 loci associated with eGFR through trans-ethnic meta-analysis

The 121 GWAS together included data from 765,348 individuals (567,460 EA, 165,726 of East Asian ancestry, 13,842 African Americans, 13,359 of South Asian ancestry, and 4,961 Hispanics, **Supplementary Table 1**). The median of the study-specific mean eGFR was 89 ml/min/1.73m² (interquartile range [IQR] 81-94), the median age was 54 years, and 50% were female. GWAS of eGFR were based on genotypes imputed using reference panels from the Haplotype Reference Consortium (HRC)[22] or the 1000 Genomes Project[23] (Methods, **Supplementary Table 2**). Following study-specific variant filtering and quality control procedures, fixed-effects inverse-variance weighted meta-analysis was conducted (Methods). There was no evidence of result inflation due to population stratification (LD Score regression intercept =1.04; λ_{GC}=1.05). After variant filtering, 8,221,591 single nucleotide polymorphisms (SNPs) were used for analysis (Methods).

We identified 308 loci containing at least one SNP associated with eGFR at genome-wide significance ($P < 5\times10^{-8}$). Loci were defined as ±500 kb around the SNP with the lowest p-value (index SNP); the extended MHC region was considered one locus (Methods). Of these loci, 200 were considered novel and 108 known because they contained an index SNP reported by previous GWAS of eGFR (**Figure 1**). Association statistics for all 308 index SNPs are shown

in **Supplementary Table 3** and the corresponding regional association plots in **Supplementary Figure 2**. The minor alleles across index SNPs showed both eGFR-decreasing and eGFR-increasing effects (**Figure 1**, inset). Together, the 308 index SNPs explained 7.1% of the eGFR variance (Methods), nearly doubling the estimate from a recent eGFR GWAS meta-analysis.[8] These SNPs explained 19.6% of the estimated genetic heritability of eGFR (h^2=39%, 95% credible interval: 32%-47%) in a large participating general-population-based pedigree-study (**Supplementary Figure 3A**; Methods).

Most of the 308 index SNPs showed homogeneous effects across studies (median I^2 =5%, IQR: 0-13%; **Supplementary Table 3**; **Fig 2A**). Only one index SNP had I^2 >50% (*UMOD-PDILT* locus, I^2=60%) as described previously[9,12] and suspected to be age-related.[24] We then investigated the heterogeneity of genetic effects that was correlated with ancestry using meta-regression[25] (Methods) and identified three index SNPs with significant ancestry-related heterogeneity at the *LINC01362*, *GATM*, and *PSD4* loci (ancestry heterogeneity p-value (p-anc-het) <0.05/308; **Figure 2A**, **Supplementary Table 3**). The index SNP at *UMOD-PDILT* did not show evidence for ancestry-related heterogeneity (p-anc-het=0.59). These results suggest similar effects across ancestries for the vast majority of the identified SNPs. Ancestry-specific results for all 308 index SNPs are reported in **Supplementary Table 4**.

Generalizability of eGFR-associated SNPs and meta-analysis of >1 million individuals

Next, we assessed whether the trans-ethnic findings were generalizable to other, independent trans-ethnic samples, using data from a GWAS meta-analysis of eGFR performed among 280,722 participants of the Million Veteran Program (MVP) from US Veterans' Administration facilities.[26] The MVP sample consisted of 20.4% African American and 79.6% EA participants (mean age 64 years, 7.4% female). For 305 of the 308 index SNPs available, generalizability was supported by various measures: first, there was almost perfect direction consistency of the eGFR effect estimates from CKDGen in MVP (302/305 SNPs, 99%; **Figure 2B**). Second, there was a very strong and significant correlation of the ranks of genetic effects (Lin's concordance correlation coefficient 0.90 (95% confidence interval (CI) 0.87; 0.92, **Figure 2B**). Third, 262/305 SNPs were at least nominally associated with eGFR (one-sided P<0.05) and 74 were genome-wide significant in MVP alone (**Supplementary Table 5**). The meta-analysis of CKDGen and MVP comprising a total of 1,038,119 individuals showed genome-wide significance for the vast majority 94% of SNPs (**Supplementary Table 5**). The SNP with the lowest p-value in this meta-analysis was rs77924615 at *UMOD-PDILT* (P=3×10^{-259}). Together, these results support the

robustness of our findings even in a setting of a smaller independent study with differences in ethnic and sex composition, trait transformation, and modeling (Methods). Formal replication analysis was not attempted because of biological and analytical differences across the CKDGen and MVP efforts.

Clinical and epidemiological landscape: association of eGFR index SNPs with blood urea nitrogen (BUN) and CKD

To evaluate whether the 308 eGFR-associated SNPs pertained to kidney function rather than creatinine metabolism, we assessed their relation to BUN, an alternative markers of kidney function that is inversely correlated with eGFR. We performed fixed-effect meta-analysis of GWAS of BUN combining 416,178 samples from 65 studies (**Supplementary Table 1**), using the same workflow as for eGFR. For the 308 SNPs, we observed the expected negative correlation between the effects on eGFR and BUN (Pearson's correlation coefficient =-0.66, 95%CI: -0.72; -0.59), with 81 SNPs significantly associated with BUN (P<0.05/308 =1.6×10^{-4}; **Figure 2C**; **Supplementary Table 6**). Overall, the GWAS meta-analysis of BUN showed no genomic inflation (λ_{GC} =1.03; LD Score regression intercept =0.98) and yielded 111 genome-wide significant loci (15 known, 96 novel, **Supplementary Figure 4**, **Supplementary Table 7**).

Despite the negative correlation in general, some of the 308 loci showed clearly distinct effects on eGFR and BUN that can plausibly be attributed to the metabolism of the respective marker. For example, the index variant in *GATM*, the gene encoding the rate-limiting enzyme in the generation of the creatinine precursor creatine, was strongly associated with eGFR but showed no association with BUN (**Figure 2C**). Likewise, genetic variants in the urea transporter encoded by *SLC14A2* were strongly associated with BUN but not with eGFR (**Supplementary Table 7**). Our comparative analysis of complementary parameters of kidney function supports that signals from most eGFR-associated index SNPs were related to the kidney's filtration function.

To assess clinical relevance of the 308 eGFR loci, we estimated their odds ratios (OR) for CKD in the general population (**Figure 1, inset**; **Supplementary Table 3**). Almost all eGFR index SNPs (99.7%) showed the expected inverse effects on CKD, and 75 SNPs were significantly associated with CKD ($P<0.05/308=1.6\times10^{-4}$). The largest effects on CKD were observed for rs77924615 at *UMOD-PDILT* (OR = 0.81, 95%CI: 0.80; 0.83), rs187355703 at *HOXD8* (OR =0.82, 95%CI: 0.77; 0.87), and rs10254101 at *PRKAG2* (OR =1.11, 95%CI: 1.09; 1.11).

Genetic correlations of eGFR with other complex traits and diseases

Lower eGFR is associated with numerous cardio-metabolic risk markers and diseases.[27] We therefore assessed the evidence for a shared genetic basis or co-regulation by evaluating genetic correlations (r_g) of the trans-ethnic GWAS meta-analysis results with those from 749 other complex traits and diseases available through LD Hub using LD score regression (Methods).[28] We observed 37 significant genetic correlations ($P<6.7\times10^{-5}$ = 0.05/749, **Supplementary Figure 5**; **Supplementary Table 8**). Other than with serum creatinine, the largest negative genetic correlations were observed for serum citrate (r_g=-0.27) and urate (r_g=-0.23), followed by anthropometric traits including lean mass and physical fitness (e.g., r_g=-0.20 with left hand grip strength). While the inverse genetic correlation with muscle mass-related traits likely reflects higher creatinine generation leading to lower creatinine-based eGFR, the genetic correlations with blood concentrations of the metabolites citrate and urate likely reflect reduced filtration function, as does the positive genetic correlation with GFR estimated from cystatin C (r_g=0.53). In summary, significant genetic correlations with eGFR reflect the two components that govern serum creatinine concentrations: its excretion via the kidney and its generation in muscle.

Functional enrichment and pathway analyses

To identify previously unknown molecular mechanisms and tissues of importance for kidney function, we assessed the enrichment of the eGFR genetic associations using tissue-specific gene expression, regulatory annotations, and gene sets and pathways (Methods).

First, we used all eGFR-associated SNPs with $P<5\times10^{-8}$ to explore enriched pathways, tissues and cell types based on gene expression data using DEPICT.[29] We identified 16 significantly enriched physiological systems, cell types and tissues highlighting several aspects:

the strongest enrichment was observed for urogenital and renal physiological systems and tissues (kidney, kidney cortex, and urinary tract; false discovery rate (FDR) <0.05; **Supplementary Figure 6A and B**), which can be considered proof-of-concept of kidney as the primary target organ. We additionally found significant enrichment for mucous membrane, respiratory mucosa, nasal mucosa, and nose (enrichment p-values from 3.1×10^{-4} to 1.2×10^{-3}), possibly reflecting epithelial cell processes including transport mechanisms that are shared with the kidney. Pathway and gene set enrichment analysis identified three highly correlated and strongly associated meta gene sets ($P<1\times10^{-6}$, FDR<0.05), including some relevant to kidney such as polyuria, dilated renal tubules, and expanded mesangial matrix, as well as signaling and transcription, and energy metabolism (**Supplementary Figure 6C**).

Second, we used the genome-wide eGFR summary statistics to identify cell-type groups with evidence for enriched heritability based on data from diverse, cell-type specific functional genomic elements using stratified LD Score regression.[30] Across 10 evaluated cell type groups, the strongest and most significant enrichment was observed for kidney (13.2-fold), followed by liver (7.3-fold) and adrenal/pancreas (5.7-fold enrichment; **Supplementary Table 9**). This analysis based on regulatory marks confirms the importance of the kidney as a target organ.

Lastly, we took a complementary approach to assess enrichment of eGFR-associated variants in genes resulting in kidney phenotypes in genetically manipulated mice.[31] We selected all genes that - when genetically manipulated - cause abnormal GFR (n=24), abnormal kidney physiology (n=453), or abnormal kidney morphology (n=764). Human orthologs for these genes were interrogated in the eGFR summary statistics for the presence of significant associations (Methods). We identified associations in 10 genes causing abnormal GFR in mice (enrichment p-value=8.9×10^{-4}), 55 causing abnormal kidney physiology (enrichment p-value=1.1×10^{-4}) and 96 causing abnormal kidney morphology (enrichment p-value=1.8×10^{-5}; **Figure 3**, Methods). Of these, 25 genes were novel, i.e. these genes had not previously been shown to contain SNPs associated with eGFR at genome-wide significance or map near known loci (**Supplementary Table 10**). These genes therefore represent additional novel eGFR-associated candidates in humans. The existing mouse models may pave the way for experimental confirmation of these findings.

Statistical fine-mapping and second signal analysis in EA individuals

To effectively fine-map loci using summary statistics, the LD reference panel needs to be ancestry-matched to the GWAS population and the sample size of the LD reference panel should scale with that of the GWAS.[32] Accordingly, conditional and fine-mapping analyses were carried out for the 256 loci that were genome-wide significantly associated with eGFR among EA participants (**Supplementary Table 11**), using LD information from 15,000 randomly selected EA individuals from the UK Biobank (Methods). Neighboring loci whose index SNPs were correlated ($r^2 \geq 0.2$) were combined, resulting in 212 non-overlapping regions. Among these regions, 277 independent genome-wide significant SNPs were identified using a forward selection approach (Methods, **Supplementary Table 12**). Together, the 277 variants explained 6.82% of the eGFR variance and 23.2% of eGFR genetic heritability in EA (**Supplementary Figure 3B**).

For each of the 277 independent variants, we computed a 99% credible set, that is, a set which contains the SNP driving the association with 99% probability.[33] The median number of SNPs per credible set was 30 (IQR: 7, 74). Twenty credible sets contained only a single SNP (at *EDEM3*, *CACNA1S*, *HOXD11*, *CPS1*, *DAB2*, *SLC34A1*, *LINC01512*, *LARP4B*, *DCDC1*, *SLC25A45*, *SLC6A13*, *GATM*, *CGNL1*, *CYP1A1*, *NRG4*, *RPL3L*, *UMOD-PDILT, SLC47A1*), and two independent sets at *BCL2L14* (**Supplementary Table 12; Figure 4**), and 58 sets contained ≤5 SNPs (small credible set). These SNPs can now be prioritized for functional studies.

To systematically examine the characteristics of the SNPs in the credible sets, SNPs were annotated with respect to their functional consequence and regulatory potential. Missense SNPs with a posterior probability of >50% of driving the association and/or mapping into a small credible set are of particular interest because they directly implicate the affected gene. Such missense SNPs were identified in 11 genes (*SLC47A1*, *RPL3L*, *SLC25A45*, *CACNA1S*, *EDEM3*, *CPS1*, *KLHDC7A*, *PPM1J*, *CERS2*, *C9,* and *SLC22A2;* **Supplementary Table 13**). Most of these variants had CADD scores >15 (**Figure 4A**, **Table 1**), a cutoff used to indicate deleteriousness.[34] As summarized in **Table 1**, several of the identified genes are plausible biological candidates. For example, the missense p.Ala465Val SNP in *SLC47A1* (posterior probability >99%) encodes an amino acid change of the encoded multidrug and toxin extrusion protein (MATE1). This transport protein is responsible for the secretion of cationic drugs, toxins and internal metabolites including creatinine across brush border membranes including kidney proximal tubules. The fact that MATE1 knockout mice have higher blood levels of both creatinine and BUN[35] argues against a sole effect on creatinine transport. Altered ability to

excrete toxic compounds via kidney tubular cells may also be the molecular mechanisms underlying the association signal at another fine-mapped missense SNP, p.Ser270Ala in *SLC22A2* (**Table 1**).

Emerging experimental evidence provides molecular mechanisms by which regulatory variants identified from GWAS exert their effects.[36] To evaluate whether small credible set SNPs may have regulatory potential in the kidney, we annotated them to regions of open chromatin identified from primary micro-dissected human tubular and glomerular cells (GEO accession number GSE115961), as well as from publicly available kidney cells types (ENCODE and Roadmaps Projects; Methods). We identified 63 SNPs in 39 credible sets that mapped into one of these annotations and may thus represent causal regulatory variants (**Supplementary Table 13**). A finding of particular interest was intronic rs77924615 in *PDILT*. This SNP had a posterior probability of >99% of driving the association signal at the well-established *UMOD* locus. It mapped into open chromatin in all evaluated resources (native kidney cells, ENCODE and Roadmap kidney cell types), implicating rs77924615 as a candidate causal regulatory variant (**Figure 4B**) associated with differential expression of uromodulin, the product of the neighboring *UMOD* gene.

Gene prioritization via gene expression co-localization analysis

A complementary approach to highlight target genes in associated loci is to systematically evaluate co-localization of the genetic associations with phenotype and gene expression (expression quantitative trait locus, eQTL) in *cis*. We performed co-localization analyses for each eGFR-associated locus with gene expression across 46 tissues including kidney glomerular and tubulo-interstitial compartments (Methods). A high posterior probability for co-localization of eGFR and eQTL signals (posterior probability of H4 >80%) in at least one kidney tissue was observed for 20 transcripts (**Figure 5**), pointing towards a shared underlying SNP and implicating the gene encoding for the co-localized transcript as the locus' effector gene(s). Among the implicated genes, there were novel candidates that could be fine-mapped to small credible sets such as *KLHDC7A*, and several biologically plausible ones. For example, results suggest that *KNG1* (linked to kidney function via the regulation of the renin-angiotensin-aldosterone system[37]) *FGF5* (effects on blood pressure,[38]) and *MLLT3* (regulation of renal reabsorption of salt[39]) are the effector genes in the respective eGFR-associated loci (**Supplementary Table 14**). Interestingly, co-localization with gene expression in the kidney supports *GALNTL5* as the causal gene at the *PRKAG2* locus, one of the earliest and strongest

signals detected in GWAS of eGFR.[12] *GALNTL5* encodes for Polypeptide N-Acetylgalactosaminyltransferase Like 5, which has been linked to sperm development but was also found expressed in native rat kidney collecting ducts.[40]

In the tubulo-interstitial compartment, the strongest and inverse change in gene expression with lower eGFR was observed for *UMOD* (**Figure 5**). This is consistent with previous reports, in which alleles associated with lower eGFR at *UMOD*[41] were associated with higher urinary uromodulin concentrations.[42] The lead SNP at this locus was rs77924615, highlighted above as the candidate causal regulatory variant mapping into the intron of *PDILT*, the neighboring gene upstream of *UMOD*. This observation leads credence to *UMOD* as the causal gene at this locus and to rs77924615 as a regulatory SNP that may map into an upstream enhancer element. It also illustrates the value of studying gene expression in diverse tissues, as this kidney-specific co-localization would have been missed had target tissue not been studied.

A further interesting finding from studying gene expression in diverse tissues was that, for some genes, lower eGFR was associated with higher transcript levels in some tissues and lower transcript levels in others, as in the case of *SH3YL1*. For other genes, such as *METTL10,* the direction of change in transcript levels was consistent across tissues (**Figure 5**). Results across all evaluated tissues with evidence for at least one co-localized eQTL, including those without co-localization in kidney, highlight additional tissue-shared and tissue-specific signals (**Supplementary Figure 7).**[43,44]

Trans-eQTL annotation of the index SNPs was only performed for whole blood and peripheral blood mononuclear cells, for which eQTL datasets with large sample size were available (Methods). Based on the analysis of 5 non-overlapping EA genome-wide eQTL studies (sample size range 1469 - 6645, **Supplementary Table 15**), we identified a reproducible link of rs17696736 (12q24.12) with both the calcium-binding protein gene *S100A10* (1q21.3) and *STAT1* (2q32.2). *S100A10* encodes a subunit of annexin A2, which co-localizes with *PLA2R* at the cell surface and in extracellular vesicles from podocytes.[45] Inhibition of *STAT1* has been reported to protect from glomerular mesangial cell senescence[46] and to ameliorate renal oxidative stress[47] (**Supplementary Table 16**).

Co-localization with protein levels supports UMOD as a target gene

The *UMOD* locus is of particular clinical interest for CKD research:[19] rare *UMOD* mutations cause autosomal-dominant tubulo-interstitial kidney disease[48] and common variants at *UMOD* give rise to the strongest signal in GWAS of eGFR and CKD.[14] We therefore investigated this locus in further detail: conditional analyses based on the EA-specific summary statistics indicated the presence of two independent variants at this locus (**Figure 6A**), with rs77924615 mapping into upstream *PDILT*, and rs34882080 mapping into an intron of *UMOD*. Association results for the urinary uromodulin-to-creatinine ratio (UUCR) in one of the participating cohorts gave a similar appearance (**Figure 6B**) with rs34262842 (r^2 with rs34882080 0.93) as the lead variant. Co-localization of the associations with eGFR and with UUCR was evaluated separately for the two independent signals, the one with lead variants in *UMOD* (**Figure 6C**) and the one represented by rs77924615 in *PDILT* (**Figure 6D**). In both regions, there was a high probability of a shared underlying variant (posterior probability of H4 =0.97 and 0.96, respectively), further supporting rs77924615 as a causal regulatory variant and *UMOD* as its effector gene.

Discussion

This trans-ethnic study represents a 5-fold increase in sample size compared to previous GWAS meta-analyses of eGFR and identified 308 eGFR-associated loci, with 200 reported here for the first time. The index SNPs at these loci explain almost twice as much eGFR variance as previously reported.[8,9,21] By using complementary kidney function traits, we highlight loci that most likely reflect the kidney's filtration function and provide a comprehensive annotation resource. Our various enrichment approaches confirm kidney as the main target tissue of the detected SNPs, and co-localization with gene expression in kidney highlighted 20 potential target genes. Conditional analyses followed by statistical fine-mapping and annotation implicated a single potentially causal SNP in 20 independent loci, and identified 11 missense SNPs directly implicating *SLC47A1*, *RPL3L*, *SLC25A45*, *CACNA1S*, *EDEM3*, *CPS1*, *KLHDC7A*, *PPM1J*, *CERS2*, *C9*, and *SLC22A2* as effector genes. The move from association to single variant and effector gene and target tissue resolution, as illustrated by the deeper analysis of the *UMOD* locus, represents an important advance and a prerequisite for translational research.

Most previous meta-analyses of GWAS of eGFR were limited to a single ancestry group[7] and did not prioritize causal variants or effector genes in associated loci. While being underpowered to uncover novel loci, one previous trans-ethnic study employed statistical fine-

mapping and resolved one signal to a single variant,[18] rs77924615 at *UMOD-PDILT*, also identified in our study. At this locus, we could further characterize the link between the causal variant at *PDILT*, *UMOD* expression in the target tissue, and UMOD protein levels. This represents a significant advancement over ten year of eGFR GWAS[14] and highlights the potential of the generated resources.

Complementary approaches including enrichment analyses based on gene expression, regulatory annotations, and gene sets and pathways highlight the kidney as the most important target organ, enabling an expansion from pure genetic associations to tissue-specific functional annotation and expression analysis. However, relatively few kidney-specific experimental datasets are publicly available as compared to other organs and tissues. For example, the kidney is not well represented in the GTEx Project and not included in its tissue-specific eQTL datasets.[43] We were able to address this issue by using a recently published eQTL dataset from glomerular and tubulo-interstitial portions of micro-dissected human kidney biopsies,[49] kidney-specific regulatory information from the ENCODE and Roadmap resources, and by obtaining regulatory information from primary cells of micro-dissected human glomerular and tubulo-interstitial kidney portions.

Functional follow-up studies of potentially causal variants will benefit from prioritized loci that show clear evidence supporting one or few variants driving the association signal. Our statistical fine-mapping workflow allowed us to prioritize such variants at single-variant resolution for 20 loci, and down to a set of ≤5 SNPs for 38 additional loci. For example, the OCT2 protein encoded by the prioritized *SLC22A2* gene is known to transport several cationic drugs such as metoprolol, cisplatin, metformin and cimetidine across the basolateral membrane of renal tubular cells.[50] The p.Ser270Ala SNP prioritized by our workflow is a known pharmacogenomic variant altering the transport of these drugs and their side-effects such as cisplatin-induced nephrotoxicity.[51] Some of these drugs are commonly prescribed to CKD patients, which may be of relevance given their already reduced eGFR. Along the same lines, the prioritized p.Ala465Val SNP in *SLC47A1* that encodes the transporter MATE1 protein may affect the ability to secrete drugs and other toxins from proximal tubular cells into the urine[52] and hence alter CKD risk.

Strengths of this project include the large trans-ethnic sample size with dense genotype imputation, a standardized and automated phenotype generation and quality control workflow, and advanced and comprehensive downstream bioinformatics analyses to prioritize causal genes and SNPs across tissues. Some limitations warrant mention. Non-European populations

were still underrepresented in our study. This highlights the potential of future trans-ethnic efforts with trans-ethnic fine-mapping analyses once larger reference panels to estimate population-specific LD become available. We used GFR estimated from serum creatinine, as done in clinical practice and observational studies, because direct measurement of kidney function is invasive, time-consuming, and burdensome. We carefully calibrated creatinine across studies, used state-of-the-art estimating equations, and distributed a centrally generated and automated script for GFR phenotype computation that all participating studies used. We also evaluated genetic associations with a complementary marker of kidney function, BUN. In addition, the generated genome-wide BUN summary statistics represent a useful resource for other studies in the field that evaluate only one kidney function biomarker, typically eGFR, or for researcher interested in organ-specific functions. Our analysis focused on SNPs present in the majority of the participating studies in order to favor the identification of signals that are broadly representative and generalizable. This choice might have limited our ability to uncover novel low-frequency or population-specific variants, which represents a complementary avenue of research. Other limitations apply to GWAS efforts in general: although co-localization with gene expression can help prioritize effector genes, these associations are based on measures from a single time point and hence cannot answer whether changes in gene expression precede changes in kidney function or occur as a consequence.

In summary, we identify and characterize 308 loci associated with eGFR and prioritize potential effector genes, driver variants and target tissues. These findings will fuel functional studies and advance the understanding of kidney function biology, a prerequisite to develop novel therapies to reduce the burden of CKD.

Online Methods

Overview

We set up a collaborative meta-analysis based on a distributive data model and quality control (QC) procedures. To maximize the level of standardization of generated phenotypes across studies, an analysis plan and a command line script (https://github.com/genepi-freiburg/ckdgen-pheno) were created centrally and provided to all participating studies, which were mostly population-based (**Supplementary Table 1**). Instructions for data processing, analysis and troubleshooting were distributed to all studies via a Wiki system (https://ckdgen.eurac.edu/mediawiki/index.php/CKDGen_Round_4_EPACTS_analysis_plan). Automatically generated summary files were uploaded centrally for phenotype quality approval of the generated phenotypes. GWAS were then run within each study and uploaded centrally. GWAS QC was performed using GWAtoolbox[53] and custom scripts to assess ancestry-matching allele frequencies and fix variant positions. All studies had their own research protocols approved by the respective local ethics committees. All participants in all studies provided written informed consent.

Phenotype definition

Each study measured serum creatinine and blood urea nitrogen (BUN) as described in **Supplementary Table 1**. When measured with a Jaffé assay before 2009, serum creatinine values were calibrated by multiplying by 0.95.[54] In studies of >18 year-old adults, eGFR was estimated with the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation,[55] using the R software package 'nephro'.[56] In studies of ≤18 year-old subjects, eGFR was estimated with the Schwartz formula[57]. eGFR values were winsorized at 15 and 200 ml/min/1.73 m^2. CKD was defined as eGFR <60 ml/ min/1.73 m^2. BUN was derived for studies that reported blood urea measurements by multiplication by 2.8, and units were aligned to mg/dl across cohorts. All steps occurred in the central phenotyping script.

Genotyping and genotype imputation

Genotype imputation was conducted based on the Haplotype Reference Consortium (HRC) version 1.1 or the 1000 Genomes Project phase 3 version 5 (1000Gp3v5) ALL or phase 1 version 3 (1000Gp1v3) ALL panels. Imputed variants were coded as allelic dosages

1 accompanied by the corresponding imputation quality (IQ) scores (ImputeV2 info score,
2 MACH/minimac RSQ, or as applicable), and annotated on the NCBI b37 (hg19) reference build.
3 Study-specific genotyping arrays as well as haplotype phasing and genotype imputation
4 methods are described in **Supplementary Table 2**.

5

6 *Genome-wide association studies* (GWAS)

7 In each study, eGFR residuals were derived from sex- and age-adjusted linear regression
8 models fitted to log(eGFR) and BUN using the centrally distributed script. In the subsequent
9 GWAS, residuals were regressed on SNP dosage levels, assuming an additive genetic model.
10 Study-specific features, such as study site, genetic principal components (PCs), or relatedness,
11 were accounted for in the study-specific models as appropriate (**Supplementary Table 2**).
12 Logistic regression models were fitted for CKD.

13

14 *Trans-ethnic GWAS meta-analysis*

15 For eGFR, 121 GWAS summary statistics files were contributed across studies. After QC, the
16 total samples size was 765,348 (567,460 individuals of European ancestry [EA], 165,726 of
17 East Asian ancestry, 13,842 African Americans, 13,359 of South Asian ancestry, and 4961
18 Hispanics; **Supplementary Table 1**). For CKD, 60 GWAS summary files were contributed,
19 totaling a post-QC samples size of 625,219 (64,164 cases). For BUN, 65 GWAS summary files
20 were contributed, totaling a post-QC samples size of 416,178 (see **Supplementary Table 1** for
21 ancestry-specific details of CKD and BUN).

22 Before meta-analysis, study-specific GWAS files were filtered to retain only variants with
23 IQ score >0.6 and minor allele count (MAC) >10. Within study, we estimated the genomic
24 control (GC) factor λ_{GC} and applied GC correction when λ_{GC} was >1. Fixed effects inverse-
25 variance weighted meta-analysis was performed using METAL,[58] which was adapted to
26 increase effects and standard errors precision (seven decimal places instead of four).

27 After meta-analysis of 43,994,957 SNPs, we retained only variants that were present in
28 ≥50% of the GWAS data files and had a total MAC of ≥400. Across ancestries, this yielded
29 8,221,591 variants for eGFR (8,834,748 in EA), 8,176,554 variants for BUN (8,358,347 in EA),
30 and 9,585,923 variants for CKD. Post-meta-analysis GC correction was not applied because the
31 LD Score regression intercept was close to 1 in all analyses of eGFR, BUN, and CKD.[59] The

1 genome-wide significance level was set at 5×10^{-8}. Between-study heterogeneity was assessed
2 using the I^2 statistic.[60] For CKD, variants with an $I^2 \geq 95\%$ were removed to avoid exaggerated
3 influence of single large studies. Variants were assigned to loci by selecting the SNP with the
4 lowest p-value genome-wide as the index SNP, defining the corresponding locus as the 1 Mb-
5 segment centered on the index SNP, and repeating the procedure until no further genome-wide
6 significant SNPs remained. A locus was considered novel if it did not contain any variant
7 identified by previous GWAS of eGFR.

8

9 *Meta-regression analysis of trans-ethnic GWAS*

10 For eGFR, we evaluated heterogeneity attributable to ancestry using quality-controlled study-
11 specific GWAS files and the software Meta-Regression of Multi-Ethnic Genetic Association (MR-
12 MEGA v0.1.2.[25]). Meta-regression models included three axes of genetic variation. GC
13 correction was applied to the meta-regression results. For the 308 genome-wide significant
14 index SNPs from the trans-ethnic GWAS meta-analysis, we tested ancestry-related
15 heterogeneity of allelic effects at a significance level of 0.05/308 = 1.6×10^{-4} (indicating the
16 corresponding p-value as p-anc-het).

17

18 *Proportion of phenotypic variance explained and genetic heritability analysis*

19 The proportion of phenotypic variance explained by the index SNPs was estimated as
20 $\beta^2\left(\frac{2p(1-p)}{var}\right)$, with β being the SNP effect, *p* the effect allele frequency, and *var* the variance of
21 the sex- and age-adjusted log(eGFR) residuals (assumed as 0.016 based on data from 11,827
22 EA participants of the population-based ARIC study).[8] The proportion of variance explained by
23 independent genome-wide significant index SNPs was estimated using the GCTA COJO Slct
24 analysis (see fine-mapping section below).

25 Genetic heritability of age- and sex-adjusted log(eGFR) was estimated using the R
26 package 'MCMCglmm'[61] in the Cooperative Health Research In South Tyrol (CHRIS) study,[62] a
27 participating pedigree-based study of EA individuals (186 up-to-5 generation pedigrees, totaling
28 4373 subjects).[63] We fitted two models, with and without the inclusion of the identified index
29 variants (304/308 and 277/277 from the transethnic and EA analyses, respectively), running
30 1,000,000 MCMC iterations (*burn in* = 500,000) based on previously described settings.[63]

1 *Comparison with results from the Million Veteran's Program (MVP) analysis*

2 To understand the robustness and generalizability of the eGFR-associated SNPs identified in
3 the CKDGen Consortium, we interrogated the effect estimates of the 308 trans-ethnic index
4 SNPs in a GWAS from an independent, large, trans-ethnic study, the Million Veteran Program
5 (MVP).[26] Briefly, the MVP study participants were recruited across 63 U.S. Veteran's
6 Administration (VA) medical facilities. Written informed consent was obtained and all documents
7 and protocols were approved by the VA Central Institutional Review Board. DNA was genotyped
8 using a customized Affymetrix Axiom Biobank Array chip with additional content added to
9 provide coverage of African and Hispanic haplotypes, as well as markers for common diseases
10 in the VA population. After QC, genotype were pre-phased using EAGLE version 2[64] and
11 imputed based on the 1000Gp3v5 reference panel using minimac3.[65] Genotype PCs were
12 estimated using FlashPCA.[66] Serum creatinine was assessed up to one year prior to MVP
13 enrollment using isotope dilution mass spectrometry. eGFR was estimated using the CKD-EPI
14 equation[55] after excluding subjects on dialysis, transplant patients, amputees, individuals on HIV
15 medications, and those with creatinine values of <0.4 mg/dl. Diabetes was defined as use of
16 anti-diabetic medications or by assignment of an International Classification of Diseases 9 (ICD-
17 9) code for diabetes during the baseline period. Hypertension was defined as having an ICD-9
18 code for hypertension, being on antihypertensive drug or having ≥2 measures of systolic or
19 diastolic blood pressure >140 mmHg or >90 mmHg, respectively.

20 GWAS of eGFR on SNP dosage levels were performed by fitting linear regression
21 models adjusted for age at creatinine measurement, age^2, sex, body mass index, and the first
22 10 genetic PCs, using SNPTEST version 2.5.4-beta.[67] All GWAS were stratified by self-reported
23 ethnicity (79.6% White non-Hispanic and 20.4% Black non-Hispanic), diabetes, and
24 hypertension status. Results were combined across strata using fixed effects inverse-variance
25 weighted meta-analysis in METAL.[58] This analysis encompassed a total of 280,722 subjects
26 across all strata. Of the 308 trans-ethnic eGFR index variants, 305 variants or their proxies
27 (minimum $r^2 \geq 0.8$, proxies selected by maximum r^2 and, in case of ties, by minimum distance)
28 were available in the MVP GWAS. CKDGen and MVP trans-ethnic meta-analysis results were
29 pooled via sample size weighted meta-analysis of z-scores using METAL.[58]

1 *Genome-wide genetic correlations with other complex traits and diseases*

2 Genome-wide genetic correlation analysis was carried out to investigate evidence of co-
3 regulation or shared genetic bases between eGFR and other complex traits and diseases, both
4 known and not known to correlate with eGFR. We estimated pairwise genetic correlation
5 coefficients (r_g) between the results of our trans-ethnic meta-analysis of eGFR and each of 749
6 pre-computed and publicly available GWAS summary statistics of complex traits and diseases
7 available through LD Hub version 1.9.0 using LD Score regression.[28] An overview of the
8 sources of these summary statistics and their corresponding sample sizes is available at
9 http://ldsc.broadinstitute.org. Statistical significance was assessed at the Bonferroni corrected
10 level of 6.7×10^{-5} (=0.05/749).

11

12 *Functional enrichment: pathway and tissue enrichment analysis*

13 We used DEPICT version 1 release 194 to perform Data-Driven Expression Prioritized
14 Integration for Complex Traits analysis,[29] including pathway/gene-set enrichment and tissue/cell
15 type analyses as described previously.[8,9] All 14,461 gene sets were reconstituted by identifying
16 genes that were transcriptionally co-regulated with other genes in a panel of 77,840 gene
17 expression microarrays,[68] from mouse knock-out studies, and molecular pathways from protein-
18 protein interaction screening. In the tissues and cell type enrichment analysis, we tested
19 whether genes in associated regions were highly expressed in 209 MeSH annotation categories
20 for 37,427 microarrays on the Affymetrix U133 Plus 2.0 Array platform. We included all variants
21 associated with eGFR at a p-value of $<5\times10^{-8}$ in the trans-ethnic meta-analysis. Independent
22 variant clumping was performed using Plink 1.9[69] with 500 kb flanking regions and $r^2>0.01$ in the
23 1000Gp1v3 dataset. After excluding the MHC region, DEPICT was run with 500 repetitions to
24 estimate the FDR and 5000 permutations to compute p-values adjusted for gene length by
25 using 500 null GWAS. All significant gene sets were merged into meta gene sets by running an
26 affinity propagation algorithm[70] implemented in the Python 'scikit-learn' package (http://scikit-
27 learn.org/). The resulting network was visualized using Cytoscape (http://cytoscape.org/).

28

29 *Enrichment of heritability by cell type group*

30 We used stratified LD Score regression as a complementary method to investigate important
31 tissues and cell types based on the trans-ethnic eGFR meta-analysis results. Heritability
32 enrichment in 10 cell type groups was assessed using the default options of stratified LD Score

1 regression described previously.[30] The 10 cell type groups were collapsed from 220 cell-type
2 specific regulatory annotations for the four histone marks H3K4me1, H3K4me3, H3K9ac, and
3 H3K27ac. The enrichment of a cell type category was defined as the proportion of SNP
4 heritability in that group divided by the proportion of SNPs in the same cell type group.

5

6 *Identification of variants in genes causing kidney phenotypes in mice*

7 A nested candidate gene analysis was performed using GenToS[71] to identify additional genetic
8 associations that were missing conventional genome-wide statistical significance. Candidate
9 genes causing kidney phenotypes in mice upon manipulation were selected using the
10 comprehensive Mouse Genome Informatics (MGI) phenotype ontology in September 2017
11 (abnormal renal glomerular filtration rate [MP:0002847]; abnormal kidney morphology
12 [MP:0002135]; abnormal kidney physiology [MP:0002136]). The human orthologs of these
13 genes were obtained using the Human-Mouse: Disease Connection webtool
14 (http://www.informatics.jax.org/humanDisease.html). Genes with no human orthologs were
15 removed. Statistical significance was defined as a Bonferroni correction of a type I error level of
16 0.05 for the number of independent common SNPs across all genes in each of the three
17 candidate gene lists plus their flanking regions, based on an ancestry-matched reference
18 population. In a next step, the GWAS meta-analysis summary statistics for eGFR were queried
19 for significantly associated SNPs mapping into the selected candidate genes. Enrichment of
20 significant genetic associations in genes within each candidate list was computed from the
21 complementary cumulative binomial distribution.[71] GenToS was used with default parameters
22 on each of the three candidate gene lists, using the 1000 Genomes phase 3 release 2 ALL
23 dataset as reference.

24

25 *Identification of independent variants in the EA meta-analysis*

26 To identify additional, independent eGFR-associated variants within the identified loci,
27 approximate conditional analyses were carried out that incorporated LD information from an
28 ancestry-matched reference population. We used the genome-wide eGFR summary statistics
29 from the EA meta-analysis as input, because an LD reference sample scaled to the size of our
30 meta-analysis was only available for EA individuals.[32] We randomly selected 15,000 participants
31 from the UK Biobank dataset (UKBB; dataset ID 8974). Individuals who withdrew consent and
32 those not meeting data cleaning requirements were excluded, keeping only those who passed

sex-consistency check, had ≥95% call rate, and did not represent outliers with respect to SNP heterozygosity. For each pair of individuals, the proportion of variants shared identical-by-descent (IBD) was computed using PLINK.[72] From pairs with IBD coefficient ≥0.1875 we retained only one member. Individuals were restricted to those of EA by excluding outliers along the first two PCs from a principal component analysis seeded with the HapMap phase 3 release 2 populations as reference. The final dataset to estimate LD included 13,558 EA individuals and 16,969,363 SNPs.

The basis for statistical fine-mapping were the 256 1-Mb genome-wide significant loci identified in the EA meta-analysis, clipping at chromosome borders. Overlapping loci as well as pairs of loci whose respective index SNPs were correlated (r^2 >0.1 in the UKBB LD dataset described above) were merged. A single SNP was chosen to represent the MHC region, resulting in a final list of 212 regions prior to fine-mapping. Within each region, the GCTA COJO Slct algorithm[73] was used to identify independent variants employing a step-wise forward selection approach. We used the default collinearity cut-off of 0.9 (sensitivity analyses showing no major influence of alternative cutoff values; data not shown). We deemed an additional SNP as independently genome-wide significant if the SNPs' p-value conditional on all previously identified SNPs in the same region was $<5\times10^{-8}$.

Statistical fine-mapping and credible set generation in the EA meta-analysis

Statistical fine-mapping was carried out for each of the 212 regions. For each region containing multiple independent SNPs and for each independent SNP in such regions, approximate conditional analyses were carried out using the GCTA COJO-Cond algorithm to generate approximate conditional association statistics conditioned on the other independent SNPs in the region. Using the Wakefield's formula implemented in the R package 'gtx',[74] we derived approximate Bayes factors (ABF) from conditional estimates in regions with multiple independent SNPs and from the original estimates for regions with a single independent SNP. Given that 95% of the SNP effects on log(eGFR) fell within the -0.01 to 0.01 interval, the standard deviation prior was chosen as 0.0051 based on formula no. 8 in the original publication.[33] Sensitivity analyses showed that results were robust when higher values were used for the standard deviation prior (data not shown). For each variant within an evaluated region, the ABF obtained from the association betas and their standard errors of the marginal (single signal region) or conditional estimates (multi-signal regions) was used to calculate the posterior probability (PP) for the variant driving the association signal ("causal variant"). Ninety-

1 nine percent credible sets, representing the set of SNPs that contain the causal variant(s) with
2 99% probability, were computed by ranking variants by their PP and adding them to the set until
3 the cumulative PP was >99% in each region.

4

5 *Variant annotation*

6 Functional annotation of variants mapping into credible sets was performed by querying the
7 SNiPA database version 3.2 (March 2017),[75] based on the 1000Gp3v5 and Ensembl version 87
8 datasets. SNiPA was also used to derive the Combined Annotation Dependent Depletion
9 (CADD) PHRED-like score,[76] based on CADD version 1.3. The Ensembl VEP tool[77] was used
10 for SNP's primary effect prediction.

11

12 *Co-localization analysis of associations with eGFR and gene expression (cis-eQTLs)*

13 As the great majority of gene expression datasets is generated based on EA ancestry samples,
14 co-localization analysis was based on the genetic associations with eGFR in the EA sample and
15 with gene expression quantified from micro-dissected human glomerular and tubulo-interstitial
16 kidney portions from 187 individuals participating in the NEPTUNE study,[49] as well as from the
17 44 tissues included in the GTEx Project version 6p release.[43] The eQTL and GWAS effect
18 alleles were harmonized. For each locus, we identified tissue gene pairs with reported eQTL
19 data within ±100 kb of each GWAS index variant. The region for each co-localization test was
20 defined as the eQTL *cis* window defined in the underlying GTEx and NephQTL studies. We
21 used the default parameters and prior definitions set in the 'coloc.fast' function from the R
22 package 'gtx' (https://github.com/tobyjohnson/gtx), which is an adaption of Giambartolomei's
23 colocalization method.[78] The package was also used to estimate the direction of effect over the
24 credible sets as the ratio of the average PP weighted GWAS effects over the PP weighted eQTL
25 effects.

26

27 *Trans-eQTL analysis*

28 We performed *trans*-eQTL annotation through LD mapping based on the 1000Gp3v5 European
29 reference panel with an r^2 cut-off of >0.8. We limited annotation to index SNPs with a fine-
30 mapping posterior probability ≥1% in at least one fine-mapping-region. Due to expected small
31 effect sizes, only genome-wide *trans*-eQTL studies of either peripheral blood mononuclear cells

or whole blood with a sample size of ≥1000 individuals were considered, resulting in five non-overlapping studies[79-83] (**Supplementary Table 15**). For the study by Kirsten *et al*,[83] we had access to an update with larger sample size combining two non-overlapping studies (LIFE-Heart[84] and LIFE-Adult[85]) resulting in a total sample size of 6645. To improve stringency of results, we focused the analysis on inter-chromosomal *trans*-eQTLs with p-values of $<5\times10^{-8}$ reported by ≥2 studies.

Co-localization analyses with urinary uromodulin concentrations

Association between concentrations of the urinary uromodulin-to-creatinine ratio with genetic variants at the *UMOD-PDILT* locus were evaluated in the German Chronic Kidney Disease (GCKD) study.[86] Uromodulin concentrations were measured from frozen stored urine using an established ELISA assay with excellent performance as described previously.[42] Concentrations were indexed to creatinine to account for urine dilution. Genetic associations were computed using the same software and settings as for the association with eGFR (**Supplementary Table 2**). Co-localization analyses were carried out using identical software and settings as described above for the association with gene expression.

Acknowledgements

We thank Daniele Di Domizio (Eurac Research) and Jochen Knaus (University of Freiburg) for IT assistance, and Toby Johnson (GSK) for sharing his code and discussion on credible set fine-mapping and co-localization analysis. This research has been conducted using the UK Biobank Resource under Application Number 20272. Study-specific acknowledgements and funding sources are listed in the **Supplementary Material** (page 10).

Author contributions

Manuscript writing group: Matthias Wuttke, Yong Li, Man Li, Karsten Sieber, Mary Feitosa, Mathias Gorski, Adrienne Tin, Lihua Wang, Holger Kirsten, Tarunveer Ahluwalia, Kevin Ho, Iris Heid, Markus Scholz, Alexander Teumer, Anna Köttgen, Cristian Pattaro

Design of this study: Carsten A. Böger, Christian Fuchsberger, Mathias Gorski, Anna Köttgen, Andrew P. Morris, Cristian Pattaro, Alexander Teumer, Adrienne Tin, Matthias Wuttke

Management of an individual contributing study: Tarunveer S. Ahluwalia, Emanuele di Angelantonio, Shreeram Akilesh, Stephan J.L. Bakker, Ginevra Biino, Murielle Bochud, Michael Boehnke, Eric Boerwinkle, Martin H. de Borst, Hermann Brenner, Adam S. Butterworth, Carsten A. Böger, Archie Campbell, Robert J. Carroll, John C. Chambers, Daniel I. Chasman, Ching-Yu Cheng, Kaare Christensen, Renata Cifkova, Marina Ciullo, Josef Coresh, Daniele Cusi, Rob M. van Dam, John Danesh, Olivier Devuyst, Cornelia M. van Duijn, Kai-Uwe Eckardt, Georg Ehret, Michele K. Evans, Janine F. Felix, Oscar H. Franco, Barry I. Freedman, Yechiel Friedlander, Ron T. Gansevoort, He Gao, Paolo Gasparini, John Michael Gaziano, Vilmantas Giedraitis, Christian Gieger, Franco Giulianini, Alessandro De Grandi, Vilmundur Gudnason, Tamara B. Harris, Pim van der Harst, Catharina A. Hartman, Caroline Hayward, Chew-Kiat Heng, Andrew A. Hicks, Kevin Ho, Adriana Hung, M. Arfan Ikram, Olafur S. Indridason, Erik Ingelsson, Vincent W.V. Jaddoe, Jost B. Jonas, Bettina Jung, Candace M. Kammerer, Chiea Chuen Khor, Wieland Kiess, Marcus E. Kleber, Wolfgang Koenig, Jaspal S. Kooner, Holly Kramer, Florian Kronenberg, Bernhard K. Krämer, Michiaki Kubo, Johanna Kuusisto, Mika Kähönen, Antje Körner, Anna Köttgen, Terho Lehtimäki, Yong Li, Su-Chi Lim, Markus Loeffler, Ruth J.F. Loos, Susanne Lucae, Mary Ann Lukas, Patrik K.E. Magnusson, Nicholas G. Martin, Deborah Mascalzoni, Koichi Matsuda, Olle Melander, Andres Metspalu, Evgenia K. Mikaelsdottir, Yuri Milaneschi, Karen L. Mohlke, Grant W. Montgomery, Andrew P. Morris, Renée de Mutsert, Winfried März, Girish N. Nadkarni, Jeffrey O'Connell, Michelle L. O'Donoghue, Albertine J. Oldehinkel, Marju Orho-Melander, Willem H. Ouwehand, Afshin Parsa, Cristian Pattaro, Sarah A. Pendergrass, Brenda W.J.H. Penninx, Thomas Perls, Markus Perola, Mario Pirastu, Ozren Polasek, Belen Ponte, Peter P. Pramstaller, Michael A. Province, Bruce M. Psaty, Ton J. Rabelink, Olli T. Raitakari, Dermot F. Reilly, Rainer Rettig, Myriam Rheinberger, Paul M. Ridker, David J. Roberts, Peter Rossing, Igor Rudan, Charumathi Sabanayagam, Veikko Salomaa, Kai-Uwe Saum, Helena Schmidt, Reinhold Schmidt, Markus Scholz, Ben Schöttker, Xueling Sim, Harold Snieder, Nicole Soranzo, Cassandra N. Spracklen, Kari Stefansson, Konstantin Strauch, Michael Stumvoll, Gardar Sveinbjornsson, Per O. Svensson, E-Shyong Tai, Bamidele O. Tayo, Yih-Chung Tham, Joachim Thiery, Adrienne Tin, Daniela Toniolo, Johanne Tremblay, Ioanna Tzoulaki, Anke Tönjes, Peter Vollenweider, Aiko P.J. de Vries, Uwe Völker, Gerard Waeber, Lars Wallentin, Ya Xing Wang, Dawn M. Waterworth, Wen Bin Wei, Harvey White, John B. Whitfield, Sarah H. Wild, James G. Wilson, Charlene Wong, Tien Yin Wong, Matthias Wuttke, Liang Xu, Qiong Yang, Masayuki Yasuda, Weihua Zhang, Alan B. Zonderman

Critical review of manuscript: Tarunveer S. Ahluwalia, Peter Almgren, Emanuele di Angelantonio, Stephan J.L. Bakker, Nisha Bansal, Mary L. Biggs, Martin H. de Borst, Erwin P. Bottinger, Thibaud S. Boutin, Hermann Brenner, Adam S. Butterworth, Carsten A. Böger, Harry Campbell, Daniel I. Chasman, Xu Chen, Yurong Cheng, Audrey Chu, Marina Ciullo, Josef Coresh, Rob M. van Dam, Graciela Delgado, Jasmin Divers, Rajkumar Dorajoo, Kai-Uwe Eckardt, Digna Velez Edward, Todd L. Edwards, Paul Elliott, Karlhans Endlich, Michele K. Evans, Mary F. Feitosa, Janine F. Felix, Oscar H. Franco, Andre Franke, Barry I. Freedman, Yechiel Friedlander, Christian Fuchsberger, He Gao, Sahar Ghasemi, Christian Gieger, Ayush Giri, Scott D. Gordon, Mathias Gorski, Daniel F. Gudbjartsson, Pavel Hamet, Tamara B. Harris, Pim van der Harst, Catharina A. Hartman, Thomsen Hauke, Caroline Hayward, Iris M. Heid, Jacklyn N. Hellwege, Chew-Kiat Heng, Kevin Ho, Anselm Hoppmann, Wei Huang, Nina Hutri-Kähönen, Shih-Jen Hwang, Olafur S. Indridason, Erik Ingelsson, Vincent W.V. Jaddoe, Johanna Jakobsdottir, Jost B. Jonas, Peter K. Joshi, Bettina Jung, Mika Kastarinen, Shona M. Kerr, Marcus E. Kleber, Wolfgang Koenig, Aldi T. Kraja, Holly Kramer, Florian Kronenberg, Bernhard K. Krämer, Mikko Kuokkanen, Mika Kähönen, Antje Körner, Anna Köttgen, Brigitte Kühnel, Markku Laakso, Leslie A. Lange, Carl D. Langefeld, Jeannette Lee, Terho Lehtimäki, Man Li, Yong Li, Wolfgang Lieb, Lars Lind, Cecilia M. Lindgren, Markus Loeffler, Ruth J.F. Loos, Leo-Pekka Lyytikäinen, Patrik K.E. Magnusson, Anubha Mahajan, Jonathan Marten, Nicholas G. Martin, Deborah Mascalzoni, Christa Meisinger, Thomas Meitinger, Olle Melander, Evgenia K. Mikaelsdottir, Kozeta Miliku, Karen L. Mohlke, Grant W. Montgomery, Dennis O. Mook-Kanamori, Renée de Mutsert, Winfried März, Girish N. Nadkarni, Mike A. Nalls, Matthias Nauck, Kjell Nikus, Boting Ning, Ilja M. Nolte, Raymond Noordam, Teresa Nutile, Michelle L. O'Donoghue, Albertine J. Oldehinkel, Marju Orho-Melander, Nicholette D. Palmer, Runolfur Palsson, Afshin Parsa, Cristian Pattaro, Sarah A. Pendergrass, Brenda W.J.H. Penninx, Markus Perola, Ozren Polasek, Michael H. Preuss, Bram P. Prins, Bruce M. Psaty, Ton J. Rabelink, Laura M. Raffield, Olli T. Raitakari, Myriam Rheinberger, Kenneth M. Rice, Paul M. Ridker, Fernando Rivadeneira, Peter Rossing, Igor Rudan, Daniela Ruggiero, Charumathi Sabanayagam, Veikko Salomaa, Kai-Uwe Saum, Markus Scholz, Christina-Alexandra Schulz, Nicole Schupf, Ben Schöttker, Sanaz Sedaghat, Karsten B. Sieber, Albert V. Smith, Harold Snieder, Cassandra N. Spracklen, Konstantin Strauch, Gardar Sveinbjornsson, Per O. Svensson, Salman M. Tajuddin, Nicholas Y. Q. Tan, Bamidele O. Tayo, Alexander Teumer, Adrienne Tin, Johanne Tremblay, Ioanna Tzoulaki, Anke Tönjes, André G. Uitterlinden, Niek Verweij, Veronique Vitart, Suzanne

1 Vogelezang, Aiko P.J. de Vries, Uwe Völker, Melanie Waldenberger, Lars Wallentin, Dawn M. Waterworth, Harvey
2 White, John B. Whitfield, Sarah H. Wild, James G. Wilson, Matthias Wuttke, Qiong Yang, Zhi Yu, Alan B. Zonderman

3 **Statistical Methods and Analysis**: Tarunveer S. Ahluwalia, Masato Akiyama, Peter Almgren, Mary L. Biggs,
4 Ginevra Biino, Mathilde Boissel, Thibaud S. Boutin, Hermann Brenner, Marco Brumat, Carsten A. Böger, Mickaël
5 Canouil, Robert J. Carroll, Jin-Fang Chai, Daniel I. Chasman, Miao Li Chee, Xu Chen, Yurong Cheng, Audrey Chu,
6 Massimiliano Cocca, Maria Pina Concas, James P. Cook, Tanguy Corre, Abbas Dehghan, Graciela Delgado, Ayse
7 Demirkan, Jasmin Divers, Rajkumar Dorajoo, Digna Velez Edward, Todd L. Edwards, Mary F. Feitosa, Janine F.
8 Felix, Barry I. Freedman, Sandra Freitag-Wolf, Christian Fuchsberger, Sahar Ghasemi, Ayush Giri, Mathias Gorski,
9 Daniel F. Gudbjartsson, Martin Gögele, Toomas Haller, Pavel Hamet, Pim van der Harst, Thomsen Hauke, Iris M.
10 Heid, Jacklyn N. Hellwege, Edith Hofer, Anselm Hoppmann, Katrin Horn, Shih-Jen Hwang, Johanna Jakobsdottir,
11 Peter K. Joshi, Navya Shilpa Josyula, Bettina Jung, Yoichiro Kamatani, Masahiro Kanai, Chiea Chuen Khor, Holger
12 Kirsten, Marcus E. Kleber, Alena Krajcoviechova, Holly Kramer, Mikko Kuokkanen, Anna Köttgen, Brigitte Kühnel,
13 Leslie A. Lange, Carl D. Langefeld, Man Li, Yong Li, Jianjun Liu, Jun Liu, Leo-Pekka Lyytikäinen, Anubha Mahajan,
14 Jonathan Marten, Jade Martins, Kozeta Miliku, Pashupati P. Mishra, Nina Mononen, Andrew P. Morris, Peter J. van
15 der Most, Winfried März, Mike A. Nalls, Matthias Nauck, Boting Ning, Damia Noce, Ilja M. Nolte, Raymond Noordam,
16 Teresa Nutile, Yukinori Okada, Cristian Pattaro, Sarah A. Pendergrass, Nicola Pirastu, Michael H. Preuss, Bram P.
17 Prins, Laura M. Raffield, Myriam Rheinberger, Kenneth M. Rice, Fernando Rivadeneira, Federica Rizzi, Rico Rueedi,
18 Kathleen A. Ryan, Yasaman Saba, Erika Salvi, Markus Scholz, Christina-Alexandra Schulz, Sanaz Sedaghat, Yuan
19 Shi, Karsten B. Sieber, Xueling Sim, Albert V. Smith, Cassandra N. Spracklen, Heather M. Stringham, Gardar
20 Sveinbjornsson, Silke Szymczak, Salman M. Tajuddin, Bamidele O. Tayo, Alexander Teumer, Chris H.L. Thio,
21 Gudmar Thorleifsson, Johanne Tremblay, Niek Verweij, Veronique Vitart, Suzanne Vogelezang, Chaolong Wang,
22 Lihua Wang, James F. Wilson, Mary K. Wojczynski, Matthias Wuttke, Yizhe Xu, Qiong Yang, Laura M. Yerges-
23 Armstrong, Weihua Zhang

24 **Subject Recruitment**: Saima Afaq, Erwin P. Bottinger, Hermann Brenner, Carsten A. Böger, Archie Campbell, Harry
25 Campbell, Miao Ling Chee, Kaare Christensen, Renata Cifkova, Marina Ciullo, Daniele Cusi, Katalin Dittrich, Michele
26 K. Evans, Valencia Foo, Barry I. Freedman, Ron T. Gansevoort, Vilmundur Gudnason, Catharina A. Hartman, Wei
27 Huang, Nina Hutri-Kähönen, Olafur S. Indridason, Marcus Ising, Vincent W.V. Jaddoe, Jost B. Jonas, Bettina Jung,
28 Candace M. Kammerer, Mika Kastarinen, Jaspal S. Kooner, Alena Krajcoviechova, Florian Kronenberg, Michiaki
29 Kubo, Mika Kähönen, Anna Köttgen, Markku Laakso, Jeannette Lee, Terho Lehtimäki, Wolfgang Lieb, Lars Lind,
30 Nicholas G. Martin, Koichi Matsuda, Christa Meisinger, Andres Metspalu, Renée de Mutsert, Winfried März, Kjell
31 Nikus, Michelle L. O'Donoghue, Isleifur Olafsson, Albertine J. Oldehinkel, Sandosh Padmanabhan, Cristian Pattaro,
32 Sarah A. Pendergrass, Brenda W.J.H. Penninx, Markus Perola, Ozren Polasek, Belen Ponte, David J. Porteous,
33 Tanja Poulain, Michael A. Province, Ton J. Rabelink, Olli T. Raitakari, Myriam Rheinberger, Paul M. Ridker, Peter
34 Rossing, Igor Rudan, Daniela Ruggiero, Veikko Salomaa, Reinhold Schmidt, Blair H. Smith, Per O. Svensson,
35 Nicholas Y. Q. Tan, Andrej Teren, Yih-Chung Tham, Johanne Tremblay, Ioanna Tzoulaki, Anke Tönjes, Simona
36 Vaccargiu, Suzanne Vogelezang, Peter Vollenweider, Aiko P.J. de Vries, Gerard Waeber, Lars Wallentin, Harvey
37 White, John B. Whitfield, Sarah H. Wild, James G. Wilson, Alan B. Zonderman, Johan Ärnlöv

38 **Bioinformatics**: Tarunveer S. Ahluwalia, Shreeram Akilesh, Peter Almgren, Daniela Baptista, Sven Bergmann,
39 Adam S. Butterworth, Carsten A. Böger, Eric Campana, Robert J. Carroll, Xu Chen, Audrey Chu, Massimiliano
40 Cocca, Maria Pina Concas, Tanguy Corre, E. Warwick Daw, Frauke Degenhardt, Abbas Dehghan, Jasmin Divers,
41 Rajkumar Dorajoo, Georg Ehret, Andre Franke, He Gao, Sahar Ghasemi, Ayush Giri, Scott D. Gordon, Mathias
42 Gorski, Pavel Hamet, Thomsen Hauke, Iris M. Heid, Edith Hofer, Anselm Hoppmann, Katrin Horn, Johanna
43 Jakobsdottir, Navya Shilpa Josyula, Chiea Chuen Khor, Holger Kirsten, Marcus E. Kleber, Alena Krajcoviechova,
44 Anna Köttgen, Carl D. Langefeld, Benjamin Lehne, Man Li, Yong Li, Jianjun Liu, Leo-Pekka Lyytikäinen, Jonathan
45 Marten, Jade Martins, Yuri Milaneschi, Pashupati P. Mishra, Karen L. Mohlke, Dennis O. Mook-Kanamori, Peter J.
46 van der Most, Reedik Mägi, Winfried März, Raymond Noordam, Teresa Nutile, Sarah A. Pendergrass, Nicola Pirastu,
47 Giorgio Pistis, Anna I. Podgornaia, Michael H. Preuss, Bram P. Prins, Federica Rizzi, Rico Rueedi, Yasaman Saba,
48 Erika Salvi, Markus Scholz, Christina-Alexandra Schulz, Sanaz Sedaghat, Christian M. Shaffer, Karsten B. Sieber,
49 Albert V. Smith, Cassandra N. Spracklen, Silke Szymczak, Johanne Tremblay, Chaolong Wang, James F. Wilson,
50 Matthias Wuttke, Yizhe Xu, Laura M. Yerges-Armstrong, Zhi Yu, Weihua Zhang

51 **Interpretation of Results**: Tarunveer S. Ahluwalia, Emanuele di Angelantonio, Carsten A. Böger, Ching-Yu Cheng,
52 Katalin Dittrich, Jasmin Divers, Rajkumar Dorajoo, Karlhans Endlich, Mary F. Feitosa, Janine F. Felix, Barry I.
53 Freedman, Sahar Ghasemi, Christian Gieger, Ayush Giri, Mathias Gorski, Pavel Hamet, Pim van der Harst, Thomsen
54 Hauke, Iris M. Heid, Kevin Ho, Katrin Horn, Wei Huang, Shih-Jen Hwang, Bettina Jung, Holger Kirsten, Wolfgang
55 Koenig, Alena Krajcoviechova, Anna Köttgen, Markku Laakso, Carl D. Langefeld, Man Li, Yong Li, Patrik K.E.
56 Magnusson, Jonathan Marten, Kozeta Miliku, Karen L. Mohlke, Andrew P. Morris, Nicholette D. Palmer, Cristian
57 Pattaro, Sarah A. Pendergrass, Bram P. Prins, Dermot F. Reilly, Myriam Rheinberger, Paul M. Ridker, Markus
58 Scholz, Sanaz Sedaghat, Karsten B. Sieber, Cassandra N. Spracklen, Per O. Svensson, Bamidele O. Tayo,
59 Alexander Teumer, Adrienne Tin, Johanne Tremblay, Ioanna Tzoulaki, André G. Uitterlinden, Niek Verweij,

Veronique Vitart, Suzanne Vogelezang, Lars Wallentin, Harvey White, Matthias Wuttke, Yizhe Xu, Masayuki Yasuda, Laura M. Yerges-Armstrong

Genotyping: Najaf Amin, Daniela Baptista, Ralph Burkhardt, Adam S. Butterworth, Carsten A. Böger, Archie Campbell, Harry Campbell, Daniel I. Chasman, Ching-Yu Cheng, E. Warwick Daw, Ayse Demirkan, Rajkumar Dorajoo, Cornelia M. van Duijn, Georg Ehret, Michele K. Evans, Mary F. Feitosa, Andre Franke, Yechiel Friedlander, Christian Fuchsberger, Ron T. Gansevoort, He Gao, Scott D. Gordon, Pavel Hamet, Pim van der Harst, Thomsen Hauke, Caroline Hayward, Chew-Kiat Heng, Wei Huang, Erik Ingelsson, Chiea Chuen Khor, Marcus E. Kleber, Wolfgang Koenig, Jaspal S. Kooner, Peter Kovacs, Aldi T. Kraja, Alena Krajcoviechova, Florian Kronenberg, Michiaki Kubo, Mika Kähönen, Antje Körner, Leslie A. Lange, Terho Lehtimäki, Leo-Pekka Lyytikäinen, Patrik K.E. Magnusson, Thomas Meitinger, Olle Melander, Yuri Milaneschi, Karen L. Mohlke, Nina Mononen, Grant W. Montgomery, Dennis O. Mook-Kanamori, Andrew P. Morris, Josyf C. Mychaleckyj, Winfried März, Mike A. Nalls, Marju Orho-Melander, Sandosh Padmanabhan, Nicholette D. Palmer, Brenda W.J.H. Penninx, Markus Perola, David J. Porteous, Michael H. Preuss, Olli T. Raitakari, Dermot F. Reilly, Fernando Rivadeneira, Federica Rizzi, Jerome I. Rotter, Daniela Ruggiero, Veikko Salomaa, Erika Salvi, Blair H. Smith, Cassandra N. Spracklen, Salman M. Tajuddin, Kent Taylor, Alexander Teumer, Daniela Toniolo, Johanne Tremblay, André G. Uitterlinden, Simona Vaccargiu, Uwe Völker, Melanie Waldenberger, Chaolong Wang, Lihua Wang, Ya Xing Wang, James G. Wilson, Mary K. Wojczynski, Alan B. Zonderman, Johan Ärnlöv

Competing interests

Winfried März is employed with Synlab Services GmbH and holds shares of Synlab Holding Deutschland GmbH. Mike A. Nalls is supported by a consulting contract between Data Tecnica International LLC and the National Institute on Aging (NIA), National Institutes of Health (NIH), Bethesda, MD, USA and consults for Illumina Inc., the Michael J. Fox Foundation, and the University of California Healthcare. Oscar H Franco works in ErasmusAGE, a center for aging research across the life course funded by Nestlé Nutrition (Nestec Ltd.); Metagenics Inc.; and AXA. Karsten B Sieber, Laura Yerges-Armstrong, Dawn M Waterworth and Mary Ann Lukas are full-time employees of GlaxoSmithKline. Michelle L. O'Donoghue received grant support from GlaxoSmithKline, Merc, Eisai, AstraZeneca and Jannsen. Harvey White: Grants and non-financial support from GlaxoSmithKline, during the conduct of the study; Advisory board of AstraZeneca, grants from Sanofi Aventis, Eli Lilly and Company, NIH, Merck Sharpe & Dohm, AstraZeneca, Omthera Pharmaceuticals, Pfizer New Zealand, Intarcia Therapeutics Inc, Elsai Inc., DalGen Products and Services, outside the submitted work. Lars Wallentin: Institutonal grants from GlaxoSmithKline, AstraZeneca, BMS, Boehringer-Ingelheim, Pfizer. Merck, Roche Diagnostics. Dermot F. Reilly and Anna I. Podgornaia are employees of Merck Sharp Dohme Corp. Markus Scholz: Consultancy of and grant support from Merck Serono not related to this project. Bruce M. Psaty serves on the DSMB of a clinical trial funded by the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. John Danesh is member of the Novartis Cardiovascular and Metabolic Advisory Board, received grant support from Novartis. Adam S. Butterworth received grants from Merck, Pfizer, Novartis, Biogen and Bioverativ and personal fees from Novartis. All other authors declare no conflicts of interest.

Figure Legends

Figure 1 – Trans-ethnic GWAS meta-analysis identifies 308 loci associated with eGFR

Circos plot: Red band: $-\log_{10}(P)$ for association with eGFR, by chromosomal position. Blue line indicates genome-wide significance ($P=5\times10^{-8}$). Black gene labels indicate novel loci, blue labels known loci. Green band: Measures of heterogeneity related to the index SNPs associated with eGFR. Dot sizes are proportional to I^2 or ancestry-related heterogeneity (p-anc-het). Blue band: $-\log_{10}(P)$ for association with CKD, by chromosomal position. Red line indicates genome-wide significance ($P=5\times10^{-8}$). Radial lines mark regions with p-anc-het < 10^{-3} or I^2 >25%. Inset: Effects of all 308 index SNPs on log(eGFR) by their minor allele frequency, color-coded by the associated odds ratio (OR) of CKD (red scale for OR≤1, blue scale for OR>1). Triangles highlight SNPs that were significantly ($P<1.6\times10^{-4}$ =0.05/308) associated with CKD.

Figure 2 – Generalizability with respect to other populations and other kidney function markers

Panel A: Measures of heterogeneity for 308 eGFR-associated index SNPs. Comparison of each variant's heterogeneity quantified as I^2 from the trans-ethnic meta-analysis (Y-axis) vs. ancestry-related heterogeneity from meta-regression ($-\log_{10}$(p-anc-het), X-axis). Histograms summarize the distribution of the heterogeneity measures on both axes. SNPs with significant p-anc-het (<1.6×10^{-4} = 0.05/308) are marked in blue and labeled, SNPs with I^2 >50% are labeled. **Panel B: Comparison of genetic effect sizes between CKDGen Consortium data (X-axis) and MVP data (Y-axis).** Blue font indicates $P<1.6\times10^{-4}$ (0.05/308) in the MVP. Error bars indicate 95% CIs. Dashed line: line of best fit. Pearson's correlation coefficient: 0.92 (95% CI: 0.90; 0.94). **Panel C: Comparison of the magnitude of the effects on eGFR (X-axis) vs. BUN (Y-axis) for the 308 eGFR-associated index SNPs.** SNPs are marked in blue when $P<1.6\times10^{-4}$ (0.05/308) in the BUN analysis. Error bars indicate 95% CIs. Dashed line: line of best fit. Pearson's correlation coefficient: -0.66 (95% CI: -0.72; -0.59).

Figure 3 – Human orthologs of genes with renal phenotypes in genetically manipulated mice are enriched for association signals with eGFR

Signals in candidate genes identified based on the murine phenotypes abnormal GFR (**Panel A**), abnormal kidney physiology (**Panel B**), and abnormal kidney morphology (**Panel C**). Y-axis: $-\log_{10}(P)$ for association with eGFR in the trans-ethnic meta-analysis for the variant with the lowest p-value in each candidate gene. Dashed line indicates genome-wide significance (P=5×10^{-8}), solid gray line indicates the significance threshold for each nested candidate gene analysis (included in lower right corner in each panel, experiment-wide significance). Orange color indicates genome-wide significance, red color experiment-wide but not genome-wide significance, and blue color indicates genes with no significantly associated SNPs. Genes are labeled when reaching experiment- but not genome-wide significance; black font for genes not mapping into loci reported in the main analysis, gray font otherwise. Enrichment p-value reported for observed number of genes with association signals below the experiment-wide threshold vs. the expected number based on the complementary cumulative binomial distribution (Methods).

Figure 4 – Credible set size (X-axis) vs. variant posterior probability (Y-axis) of 4,060 variants in 212 99% credible sets by annotation

Panel A: Exonic variants. Variants are marked by triangles, with size proportional to their CADD score. Red triangles and variant labeling indicate missense variants mapping into small (≤5 SNPs) credible sets or with high individual posterior probability of driving the association signal (>0.5). **Panel B: Regulatory potential.** Symbol colors identify variants with regulatory potential as derived from DNAse hypersensitivity analysis in target tissues (Methods). Variant annotation was restricted to variants with variant posterior probability >1%; SNPs with posterior probability ≥90% contained in credible sets with ≤10 variants were labeled.

Figure 5 – Co-localization of eGFR-association signals with gene expression in kidney tissues

All eGFR loci were tested for co-localization with all eQTLs where the eQTL cis-window overlapped (±100 kb) the sentinel genetic variants. Genes with ≥1 positive co-localization (posterior probability of one common causal variant, H4, ≥0.80) in a kidney tissue are illustrated with the respective sentinel variants (Y-axis). Co-localizations across all tissues (X-axis) are

illustrated as dots, where the size of the dots indicates the posterior probability of the co-localization. Negative co-localizations (posterior probability of H4 <0.80) are marked in grey, while the positive co-localizations are color-coded based on the predicted change in expression relative to the allele associated with lower eGFR.

Figure 6 – Co-localization of independent eGFR-association signals at the *UMOD*/*PDILT* locus with urinary uromodulin concentrations supports *UMOD* as the effector gene.

Association plots: association $-\log_{10}$(p-value) (Y axis) vs. chromosomal position (X axis). Approximate conditional analyses among EA individuals support the presence of two independent eGFR-associated signals (**Panel A**). The association signal with urinary uromodulin/creatinine levels looks similar (**Panel B**). Co-localization of association with eGFR (upper sub-panel) and urinary uromodulin/creatinine levels (lower sub-panel) for the independent regions centered on *UMOD* (**Panel C**) and *PDILT* (**Panel D**) support a shared underlying variant in both regions with high posterior probability.

Table 1 – Genes implicated as causal via identification of missense variants with high probability of driving the eGFR association signal. Genes are included if they contain a missense variant with posterior probability of association of >50% or mapping into a small credible set (≤5 variants).

Gene	SNP	Credible set size	SNP PP[1]	functional consequence	CADD score[2]	DHS[3], tissue	**Brief summary of the gene's function and relevant literature** (OMIM entries are indicated as #number)
CACNA1S	rs3850625	1	1.00	p.Arg1539Cys (NP_000060.2)	34.0	-	Encodes a subunit of the slowly inactivating L-type voltage-dependent calcium channel in skeletal muscle. Reports of altered expression in kidney cancer (PMID 28781648) and after indoxyl sulfate treatment (PMID: 27550174). Rare variants can cause autosomal dominant hypokalemic periodic paralysis, type 1 (#170400) or malignant hyperthermia susceptibility (#601887). Common variation at this locus has been reported as associated with eGFR in previous GWAS (PMID: 24029420, PMID: 26831199).
CPS1	rs1047891	1	1.00	p.Thr1412Asn (NP_001116105.1)	22.1	-	Encodes a key mitochondrial enzyme of the urea cycle that catalyzes the synthesis of carbamoyl phosphate from ammonia and bicarbonate to remove excess urea. Rare mutations cause autosomal recessive carbamoylphosphate synthetase I deficiency (#237300). GWAS locus for eGFR (PMID: 20383146), serum metabolites (PMID: 23378610), and urinary glycine (PMID: 26352407), as well as for many other quantitative biomarkers. This variant has been reported to associate with hyperammonemia after valproate therapy (PMID: 23997965).
EDEM3	rs78444298	1	1.00	p.Pro746Ser (NP_079467.3)	24.6	-	The gene product accelerates the glycoprotein ER-associated degradation by proteasomes by catalyzing mannose trimming from Man8GlcNAc2 to Man7GlcNAc2 in the N-glycans. This variant has been identified by a previous exome chip association study with eGFR (PMID: 27920155).
KLHDC7A	rs11261022	7	0.71	p.Arg160Ser (NP_689588.2)	1.1	Roadmap, ENCODE kidney	Kelch Domain Containing 7A is a protein coding gene and a paralog of *KBTBD11*. No specific entry in relation to kidney disease in PubMed.
RPL3L	rs113956264	1	1.00	p.Val262Met (NP_005052.1)	27.2	-	The gene product shares sequence similarity with ribosomal protein L3. It has a tissue-specific expression pattern, with highest levels in skeletal muscle and heart.
SLC25A45	rs34400381	1	1.00	p.Arg285Cys (NP_001070709.2)	26.0	ENCODE kidney	Belongs to the SLC25 family of mitochondrial carrier proteins and is an orphan transporter. This variant has already been identified in a GWAS of symmetric dimethylarginine levels (PMID: 24159190) and in a whole-genome sequence (WGS) analysis of serum creatinine (PMID: 25082825). *SLC25A45* may play a role in biosynthesis of arginine, which is involved in the synthesis of creatine.
SLC47A1	rs111653425	1	1.00	p.Ala465Val (NP_060712.2)	24.6	-	Encodes the multidrug and toxin extrusion protein (MATE1), a transport protein responsible for the secretion of cationic drugs and creatinine across brush border membranes. This variant has already been identified in a WGS analysis of serum creatinine from Iceland (PMID: 25082825). Rare and common variants in the locus have been identified in exome chip (PMID: 27920155) and in GWAS (PMID: 20383146) studies of eGFR, respectively. MATE1 knockout (KO) mice show higher levels of serum creatinine and BUN (PMID: 19332510), arguing against a sole effect on creatinine transport and supporting an effect on kidney function.
PPM1J	rs34611728	5	0.02	p.Leu213Phe (NP_005158.5)	13.1	ENCODE kidney	This gene encodes the serine/threonine protein phosphatase. The variant has been reported in association with eGFR in an exome chip association study (PMID: 27920155).
CERS2	rs267738	5	0.46	p.Glu115Ala (NP_071358.1)	32.0/ 28.2	-	Encodes Ceramide Synthase 2, which may be involved in sphingolipid synthesis. Changes in ceramides were reported as essential in renal Madin-Darby Canine Kidney (MDCK) cell differentiation (PMID: 28515139). *CERS2* KO mice show strongly reduced ceramide levels in the kidney and develop renal parenchyma abnormalities (PMID: 19801672). This variant has been reported as associated with the rate of albuminuria increase in patients with diabetes (PMID: 25238615).
C9	rs700233	5	0.32	p.Arg5Trp (NP_001728.1)	6.6	-	Encodes a constituent of the membrane attack complex that plays a key role in the innate and adaptive immune response. Rare mutations can cause C9 deficiency (#613825). *C9* is mentioned in several kidney disease case reports, including patients with congenital factor 9 deficiency showing IgA nephropathy (PMID: 1453611).
SLC22A2	rs316019	4	0.04	p.Ser270Ala (NP_003049.2)	12.7	-	Encodes the polyspecific organic cation transporter (OCT2) that is primarily expressed in the kidney, where it mediates tubular uptake of organic compounds including creatinine from the circulation. Many publications relate *SLC22A2* to kidney function. rs316019 is a known pharmacogenomics variant associated with response to metformin and other drugs such as cisplatin. Carriers of the risk allele have a higher risk of cisplatin-induced nephrotoxicity (PMID: 19625999), indicating that this transporter is essential in excreting toxins. The locus has been reported in previous GWAS of eGFR (PMID: 20383146).

[1]PP: posterior probability. [2]CADD score: Combined Annotation Dependent Depletion (CADD) PHRED-like score (Methods); [3]DHS: DNAse Hypersensitivity Site

References

1. Eckardt, K.U. *et al.* Evolving importance of kidney disease: from subspecialty to global health burden. *Lancet* **382**, 158-69 (2013).
2. Jha, V. *et al.* Chronic kidney disease: global dimension and perspectives. *Lancet* **382**, 260-72 (2013).
3. Levin, A. *et al.* Global kidney health 2017 and beyond: a roadmap for closing gaps in care, research, and policy. *Lancet* **390**, 1888-1917 (2017).
4. Ene-Iordache, B. *et al.* Chronic kidney disease and cardiovascular risk in six regions of the world (ISN-KDDC): a cross-sectional study. *Lancet Glob Health* **4**, e307-19 (2016).
5. Inrig, J.K. *et al.* The landscape of clinical trials in nephrology: a systematic review of Clinicaltrials.gov. *Am J Kidney Dis* **63**, 771-80 (2014).
6. Wanner, C. *et al.* Empagliflozin and Progression of Kidney Disease in Type 2 Diabetes. *N Engl J Med* **375**, 323-34 (2016).
7. Wuttke, M. & Kottgen, A. Insights into kidney diseases from genome-wide association studies. *Nat Rev Nephrol* **12**, 549-62 (2016).
8. Gorski, M. *et al.* 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function. *Sci Rep* **7**, 45040 (2017).
9. Pattaro, C. *et al.* Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nat Commun* **7**, 10023 (2016).
10. Chasman, D.I. *et al.* Integration of genome-wide association studies with biological knowledge identifies six novel genes related to kidney function. *Hum Mol Genet* **21**, 5329-43 (2012).
11. Pattaro, C. *et al.* Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet* **8**, e1002584 (2012).
12. Kottgen, A. *et al.* New loci associated with kidney function and chronic kidney disease. *Nat Genet* **42**, 376-84 (2010).
13. Chambers, J.C. *et al.* Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet* **42**, 373-5 (2010).
14. Kottgen, A. *et al.* Multiple loci associated with indices of renal function and chronic kidney disease. *Nat Genet* **41**, 712-7 (2009).
15. Okada, Y. *et al.* Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat Genet* **44**, 904-9 (2012).
16. Hishida, A. *et al.* Genome-Wide Association Study of Renal Function Traits: Results from the Japan Multi-Institutional Collaborative Cohort Study. *Am J Nephrol* **47**, 304-316 (2018).
17. Lee, J. *et al.* Genome-wide association analysis identifies multiple loci associated with kidney disease-related traits in Korean populations. *PLoS One* **13**, e0194044 (2018).
18. Mahajan, A. *et al.* Trans-ethnic Fine Mapping Highlights Kidney-Function Genes Linked to Salt Sensitivity. *Am J Hum Genet* **99**, 636-646 (2016).
19. Devuyst, O. & Pattaro, C. The UMOD Locus: Insights into the Pathogenesis and Prognosis of Kidney Disease. *J Am Soc Nephrol* **29**, 713-726 (2018).
20. Yeo, N.C. *et al.* Shroom3 contributes to the maintenance of the glomerular filtration barrier integrity. *Genome Res* **25**, 57-65 (2015).
21. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* **50**, 390-400 (2018).
22. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).
23. Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).

24. Gudbjartsson, D.F. *et al.* Association of variants at UMOD with chronic kidney disease and kidney stones-role of age and comorbid diseases. *PLoS Genet* **6**, e1001039 (2010).
25. Magi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet* **26**, 3639-3650 (2017).
26. Gaziano, J.M. *et al.* Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* **70**, 214-23 (2016).
27. Go, A.S., Chertow, G.M., Fan, D., McCulloch, C.E. & Hsu, C.Y. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med* **351**, 1296-305 (2004).
28. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-41 (2015).
29. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
30. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet* **47**, 1228-35 (2015).
31. Jing, J. *et al.* Combination of mouse models and genomewide association studies highlights novel genes associated with human kidney function. *Kidney Int* **90**, 764-73 (2016).
32. Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am J Hum Genet* **101**, 539-551 (2017).
33. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* **81**, 208-27 (2007).
34. Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet* **24**, 2125-37 (2015).
35. Tsuda, M. *et al.* Targeted disruption of the multidrug and toxin extrusion 1 (mate1) gene in mice reduces renal secretion of metformin. *Mol Pharmacol* **75**, 1280-6 (2009).
36. Karczewski, K.J. & Snyder, M.P. Integrative omics for health and disease. *Nat Rev Genet* **19**, 299-310 (2018).
37. Lieb, W. *et al.* Genome-wide meta-analyses of plasma renin activity and concentration reveal association with the kininogen 1 and prekallikrein genes. *Circ Cardiovasc Genet* **8**, 131-40 (2015).
38. Newton-Cheh, C. *et al.* Genome-wide association study identifies eight loci associated with blood pressure. *Nat Genet* **41**, 666-76 (2009).
39. Zhang, W., Xia, X., Reisenauer, M.R., Hemenway, C.S. & Kone, B.C. Dot1a-AF9 complex mediates histone H3 Lys-79 hypermethylation and repression of ENaCalpha in an aldosterone-sensitive manner. *J Biol Chem* **281**, 18059-68 (2006).
40. Yu, M.J. *et al.* Large-scale quantitative LC-MS/MS analysis of detergent-resistant membrane proteins from rat renal collecting duct. *Am J Physiol Cell Physiol* **295**, C661-78 (2008).
41. Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension. *Nat Genet* **41**, 677-87 (2009).
42. Olden, M. *et al.* Common variants in UMOD associate with urinary uromodulin levels: a meta-analysis. *J Am Soc Nephrol* **25**, 1869-82 (2014).
43. Battle, A., Brown, C.D., Engelhardt, B.E. & Montgomery, S.B. Genetic effects on gene expression across human tissues. *Nature* **550**, 204-213 (2017).
44. Gamazon, E.R. *et al.* Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet* **50**, 956-967 (2018).
45. Fresquet, M. *et al.* PLA2R binds to the annexin A2-S100A10 complex in human podocytes. *Sci Rep* **7**, 6876 (2017).

46. Jiao, S., Zheng, X., Yang, X., Zhang, J. & Wang, L. Losartan inhibits STAT1 activation and protects human glomerular mesangial cells from angiotensin II induced premature senescence. *Can J Physiol Pharmacol* **90**, 89-98 (2012).
47. Lopez-Sanz, L. *et al.* SOCS1-targeted therapy ameliorates renal and vascular oxidative stress in diabetes via STAT1 and PI3K inhibition. *Lab Invest* (2018).
48. Eckardt, K.U. *et al.* Autosomal dominant tubulointerstitial kidney disease: diagnosis, classification, and management--A KDIGO consensus report. *Kidney Int* **88**, 676-83 (2015).
49. Gillies, C.E. *et al.* An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. *Am J Hum Genet* **103**, 232-244 (2018).
50. Dudley, A.J., Bleasby, K. & Brown, C.D. The organic cation transporter OCT2 mediates the uptake of beta-adrenoceptor antagonists across the apical membrane of renal LLC-PK(1) cell monolayers. *Br J Pharmacol* **131**, 71-9 (2000).
51. Filipski, K.K., Mathijssen, R.H., Mikkelsen, T.S., Schinkel, A.H. & Sparreboom, A. Contribution of organic cation transporter 2 (OCT2) to cisplatin-induced nephrotoxicity. *Clin Pharmacol Ther* **86**, 396-402 (2009).
52. Motohashi, H. & Inui, K. Organic cation transporter OCTs (SLC22) and MATEs (SLC47) in the human kidney. *AAPS J* **15**, 581-8 (2013).
53. Fuchsberger, C., Taliun, D., Pramstaller, P.P. & Pattaro, C. GWAtoolbox: an R package for fast quality control and handling of genome-wide association studies meta-analysis data. *Bioinformatics* **28**, 444-5 (2012).
54. Coresh, J. *et al.* Decline in estimated glomerular filtration rate and subsequent risk of end-stage renal disease and mortality. *JAMA* **311**, 2518-2531 (2014).
55. Levey, A.S. *et al.* A new equation to estimate glomerular filtration rate. *Ann Intern Med* **150**, 604-12 (2009).
56. Pattaro, C. *et al.* Estimating the glomerular filtration rate in the general population using different equations: effects on classification and association. *Nephron Clin Pract* **123**, 102-11 (2013).
57. Schwartz, G.J. *et al.* Improved equations estimating GFR in children with chronic kidney disease using an immunonephelometric determination of cystatin C. *Kidney Int* **82**, 445-53 (2012).
58. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
59. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
60. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**, 1539-58 (2002).
61. Hadfield, J. MCMC methods for multi-response generalized linear mixed models: the MCMC glmm R Package. *J Stat Softw* **33**, 1-22 (2010).
62. Pattaro, C. *et al.* The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. *J Transl Med* **13**, 348 (2015).
63. Noce, D. *et al.* Sequential recruitment of study participants may inflate genetic heritability estimates. *Hum Genet* **136**, 743-757 (2017).
64. Loh, P.R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443-1448 (2016).
65. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
66. Abraham, G. & Inouye, M. Fast principal component analysis of large-scale genome-wide data. *PLoS One* **9**, e93766 (2014).

67. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).
68. Fehrmann, R.S. *et al.* Gene expression analysis identifies global gene dosage sensitivity in cancer. *Nat Genet* **47**, 115-25 (2015).
69. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
70. Frey, B.J. & Dueck, D. Clustering by passing messages between data points. *Science* **315**, 972-6 (2007).
71. Hoppmann, A.S., Schlosser, P., Backofen, R., Lausch, E. & Kottgen, A. GenToS: Use of Orthologous Gene Information to Prioritize Signals from Human GWAS. *PLoS One* **11**, e0162466 (2016).
72. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
73. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
74. Wakefield, J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet Epidemiol* **33**, 79-86 (2009).
75. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmuller, G. SNiPA: an interactive, genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334-6 (2015).
76. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
77. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069-70 (2010).
78. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
79. Zeller, T. *et al.* Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PLoS One* **5**, e10693 (2010).
80. Fehrmann, R.S. *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet* **7**, e1002197 (2011).
81. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* **45**, 1238-1243 (2013).
82. Joehanes, R. *et al.* Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol* **18**, 16 (2017).
83. Kirsten, H. *et al.* Dissecting the genetics of the human transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding locidagger. *Hum Mol Genet* **24**, 4746-63 (2015).
84. Beutner, F. *et al.* Rationale and design of the Leipzig (LIFE) Heart Study: phenotyping and cardiovascular characteristics of patients with coronary artery disease. *PLoS One* **6**, e29070 (2011).
85. Loeffler, M. *et al.* The LIFE-Adult-Study: objectives and design of a population-based cohort study with 10,000 deeply phenotyped adults in Germany. *BMC Public Health* **15**, 691 (2015).
86. Eckardt, K.U. *et al.* The German Chronic Kidney Disease (GCKD) study: design and methods. *Nephrol Dial Transplant* **27**, 1454-60 (2012).