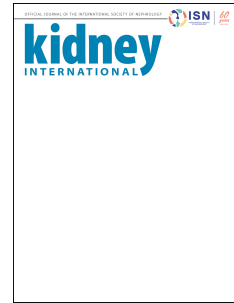# Journal Pre-proof

Genetic loci and prioritization of genes for kidney function decline derived from a meta-analysis of 62 longitudinal genome-wide association studies.

Mathias Gorski, Humaira Rasheed, Alexander Teumer, Laurent F. Thomas, Sarah E. Graham, Gardar Sveinbjornsson, Thomas W. Winkler, Felix Günther, Klaus J. Stark, Jin-Fang Chai, Bamidele O. Tayo, Matthias Wuttke, Yong Li, Adrienne Tin, Tarunveer S. Ahluwalia, Johan Ärnlöv, Bjørn Olav Åsvold, Stephan J.L. Bakker, Bernhard Banas, Nisha Bansal, Mary L. Biggs, Ginevra Biino, Michael Böhnke, Eric Boerwinkle, Erwin P. Bottinger, Hermann Brenner, Ben Brumpton, Robert J. Carroll, Layal Chaker, John Chalmers, Miao-Li Chee, Miao-Ling Chee, Ching-Yu Cheng, Audrey Y. Chu, Marina Ciullo, Massimiliano Cocca, James P. Cook, Josef Coresh, Daniele Cusi, Martin H. de Borst, Frauke Degenhardt, Kai-Uwe Eckardt, Karlhans Endlich, Michele K. Evans, Mary F. Feitosa, Andre Franke, Sandra Freitag-Wolf, Christian Fuchsberger, Piyush Gampawar, Ron T. Gansevoort, Mohsen Ghanbari, Sahar Ghasemi, Vilmantas Giedraitis, Christian Gieger, Daniel F. Gudbjartsson, Stein Hallan, Pavel Hamet, Asahi Hishida, Kevin Ho, Edith Hofer, Bernd Holleczek, Hilma Holm, Anselm Hoppmann, Katrin Horn, Nina Hutri-Kähönen, Kristian Hveem, Shih-Jen Hwang, M. Arfan Ikram, Navya Shilpa Josyula, Bettina Jung, Mika Kähönen, Irma Karabegović, Chiea-Chuen Khor, Wolfgang Koenig, Holly Kramer, Bernhard K. Krämer, Brigitte Kühnel, Johanna Kuusisto, Markku Laakso, Leslie A. Lange, Terho Lehtimäki, Man Li, Wolfgang Lieb, Lifelines cohort study, Lars Lind, Cecilia M. Lindgren, Ruth J.F. Loos, Mary Ann Lukas, Leo-Pekka Lyytikäinen, Anubha Mahajan, Pamela R. Matias-Garcia, Christa Meisinger, Thomas Meitinger, Olle Melander, Yuri Milaneschi, Pashupati P. Mishra, Nina Mononen, Andrew P. Morris, Josyf C. Mychaleckyj, Girish N. Nadkarni, Mariko Naito, Masahiro Nakatochi, Mike A. Nalls, Matthias Nauck, Kjell Nikus, Boting Ning, Ilja M. Nolte, Teresa Nutile, Michelle L. O'Donoghue, Jeffrey O'Connell, Isleifur Olafsson, Marju Orho-Melander, Afshin Parsa, Sarah A. Pendergrass, Brenda W.J. H. Penninx, Mario Pirastu, Michael H. Preuss, Bruce M. Psaty, Laura M. Raffield, Olli T. Raitakari, Myriam Rheinberger, Kenneth M. Rice, Federica Rizzi, Alexander R. Rosenkranz, Peter Rossing, Jerome I. Rotter, Daniela Ruggiero, Kathleen A. Ryan, Charumathi Sabanayagam, Erika Salvi, Helena Schmidt, Reinhold Schmidt, Markus Scholz, Ben Schöttker, Christina-Alexandra Schulz, Sanaz Sedaghat, Christian M. Shaffer, Karsten B. Sieber, Xueling Sim, Mario Sims, Harold Snieder, Kira J. Stanzick, Unnur Thorsteinsdottir, Hannah Stocker, Konstantin Strauch, Heather M. Stringham, Patrick Sulem, Silke Szymczak, Kent D. Taylor, Chris H.L. Thio, Johanne Tremblay, Simona Vaccargiu, Pim van der Harst, Peter J. van der Most, Niek Verweij, Uwe Völker, Kenji Wakai, Melanie Waldenberger, Lars Wallentin, Stefan Wallner, Judy Wang, Dawn M. Waterworth, Harvey D. White, Cristen J. Willer, Tien-Yin Wong, Mark Woodward, Qiong Yang, Laura M. Yerges-Armstrong, Martina Zimmermann, Alan B. Zonderman, Tobias Bergler, Kari Stefansson, Carsten A. Böger, Cristian Pattaro, Anna Köttgen, Florian Kronenberg, Iris M. Heid

Please cite this article as: Gorski M, Rasheed H, Teumer A, Thomas LF, Graham SE, Sveinbjornsson G, Winkler TW, Günther F, Stark KJ, Chai JF, Tayo BO, Wuttke M, Li Y, Tin A, Ahluwalia TS, Ärnlöv J, Åsvold BO, Bakker SJL, Banas B, Bansal N, Biggs ML, Biino G, Böhnke M, Boerwinkle E, Bottinger EP, Brenner H, Brumpton B, Carroll RJ, Chaker L, Chalmers J, Chee ML, Chee ML, Cheng CY, Chu AY, Ciullo M, Cocca M, Cook JP, Coresh J, Cusi D, de Borst MH, Degenhardt F, Eckardt KU, Endlich K, Evans MK, Feitosa MF, Franke A, Freitag-Wolf S, Fuchsberger C, Gampawar P, Gansevoort RT, Ghanbari M, Ghasemi S, Giedraitis V, Gieger C, Gudbjartsson DF, Hallan S, Hamet P, Hishida A, Ho K, Hofer E, Holleczek B, Holm H, Hoppmann A, Horn K, Hutri-Kähönen N, Hveem K, Hwang SJ, Ikram MA, Josyula NS, Jung B, Kähönen M, Karabegović I, Khor CC, Koenig W, Kramer H, Krämer BK, Kühnel B, Kuusisto J, Laakso M, Lange LA, Lehtimäki T, Li M, Lieb W, Lifelines cohort study, Lind L, Lindgren CM, Loos RJF, Lukas MA, Lyytikäinen LP, Mahajan A, Matias-Garcia PR, Meisinger C, Meitinger T, Melander O, Milaneschi Y, Mishra PP, Mononen N, Morris AP, Mychaleckyj JC, Nadkarni GN, Naito M, Nakatochi M, Nalls MA, Nauck M, Nikus K, Ning B, Nolte IM, Nutile T, O'Donoghue ML, O'Connell J, Olafsson I, Orho-Melander M, Parsa A, Pendergrass SA, Penninx BWJH, Pirastu M, Preuss MH, Psaty BM, Raffield LM, Raitakari OT, Rheinberger M, Rice KM, Rizzi F, Rosenkranz AR, Rossing P, Rotter JI, Ruggiero D, Ryan KA, Sabanayagam C, Salvi E, Schmidt H, Schmidt R, Scholz M, Schöttker B, Schulz CA, Sedaghat S, Shaffer CM, Sieber KB, Sim X, Sims M, Snieder H, Stanzick KJ, Thorsteinsdottir U, Stocker H, Strauch K, Stringham HM, Sulem P, Szymczak S, Taylor KD, Thio CHL, Tremblay J, Vaccargiu S, van der Harst P, van der Most PJ, Verweij N, Völker U, Wakai K, Waldenberger M, Wallentin L, Wallner S, Wang J, Waterworth DM, White HD, Willer CJ, Wong TY, Woodward M, Yang Q, Yerges-Armstrong LM, Zimmermann M, Zonderman AB, Bergler T, Stefansson K, Böger CA, Pattaro C, Köttgen A, Kronenberg F, Heid IM, Genetic loci and prioritization of genes for kidney function decline derived from a meta-analysis of 62 longitudinal genome-wide association studies., *Kidney International* (2022), doi: https://doi.org/10.1016/j.kint.2022.05.021.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Genetic loci and prioritization of genes for kidney function decline derived from a meta-analysis of 62 longitudinal genome-wide association studies.
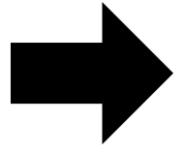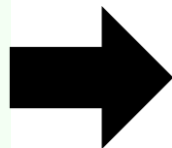
Longitudinal cohort study data

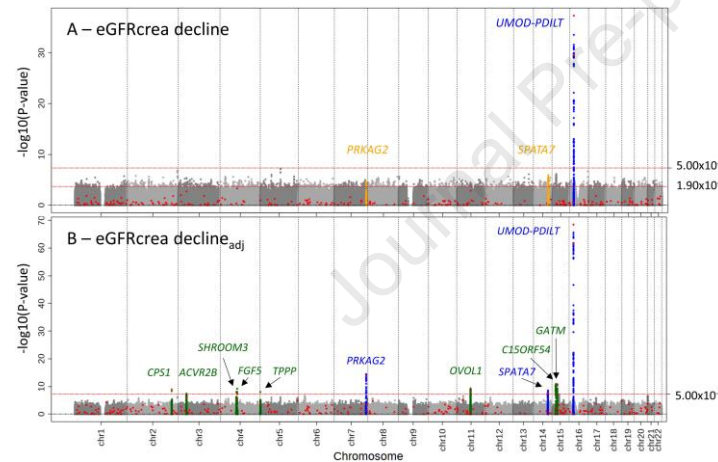>340,000 individuals with two assessments of eGFR over time

Cross-sectional study data

>350,000 further individuals with one assessment of eGFR

GWAS on eGFR decline

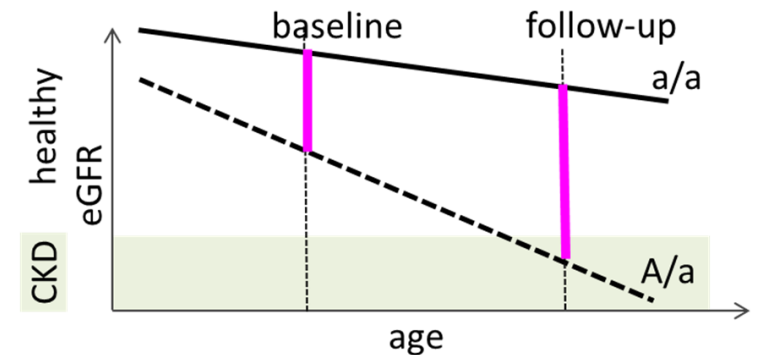Identified nine independent variants associated with eGFR-decline unadjusted for eGFR-baseline

SNP-by-age interaction on eGFR

Findings/Interpretation

UMOD   TPPP

GALTNL5   FGF5

SPATA7

CONCLUSION: We provide a large-data resource, genetic loci and prioritized genes for kidney function decline, which help inform drug development for disease progression. Results reveal important insights into the age-dependency of kidney function genetics.

Gorski et al., 2022

Visual Abstract by Iris M. Heid and Mathias Gorski

[QUERY TO AUTHOR: title and abstract rewritten by Editorial Office – not subject to change]

Genetic loci and prioritization of genes for kidney function decline derived from a meta-analysis of 62 longitudinal genome-wide association studies.

Mathias Gorski[1,2,170], Humaira Rasheed[3,4,170], Alexander Teumer[5,6,7,170], Laurent F. Thomas[3,8,9], Sarah E. Graham[10], Gardar Sveinbjornsson[11], Thomas W. Winkler[1], Felix Günther[1,12], Klaus J. Stark[1], Jin-Fang Chai[13], Bamidele O. Tayo[14], Matthias Wuttke[15,16], Yong Li[15], Adrienne Tin[17,18], Tarunveer S. Ahluwalia[19,20], Johan Ärnlöv[21,22], Bjørn Olav Åsvold[3,23], Stephan J. L. Bakker[24], Bernhard Banas[2], Nisha Bansal[25,26], Mary L. Biggs[27,28], Ginevra Biino[29], Michael Böhnke[30], Eric Boerwinkle[31], Erwin P. Bottinger[32,33], Hermann Brenner[34,35], Ben Brumpton[3,4,36], Robert J. Carroll[37], Layal Chaker[38,39], John Chalmers[40], Miao-Li Chee[41], Miao-Ling Chee[41], Ching-Yu Cheng[41,42,43], Audrey Y. Chu[44], Marina Ciullo[45,46], Massimiliano Cocca[47], James P. Cook[48], Josef Coresh[49], Daniele Cusi[50,51], Martin H. de Borst[24], Frauke Degenhardt[52], Kai-Uwe Eckardt[53,54], Karlhans Endlich[6,55], Michele K. Evans[56], Mary F Feitosa[57], Andre Franke[52], Sandra Freitag-Wolf[58], Christian Fuchsberger[59,30], Piyush Gampawar[60], Ron T. Gansevoort[24], Mohsen Ghanbari[38,61], Sahar Ghasemi[5,6], Vilmantas Giedraitis[62], Christian Gieger[63,64,65], Daniel F Gudbjartsson[11,66], Stein Hallan[8,67], Pavel Hamet[68,69,70], Asahi Hishida[71], Kevin Ho[72,73], Edith Hofer[74,75], Bernd Holleczek[34], Hilma Holm[11], Anselm Hoppmann[15], Katrin Horn[76,77], Nina Hutri-Kähönen[78,79], Kristian Hveem[3], Shih-Jen Hwang[80,81], M. Arfan Ikram[38], Navya Shilpa Josyula[82], Bettina Jung[2,83,84], Mika Kähönen[85,86], Irma Karabegović,[38] Chiea-Chuen Khor[41,87], Wolfgang Koenig[88,89,90], Holly Kramer[14,91], Bernhard K. Krämer[92], Brigitte Kühnel[63], Johanna Kuusisto[93,94], Markku Laakso[93,94], Leslie A. Lange[95], Terho Lehtimäki[96,97], Man Li[98], Wolfgang Lieb[99], Lifelines cohort study[100], Lars Lind[101], Cecilia M. Lindgren[102,103,104,105,106], Ruth J. F. Loos[32,107], Mary Ann Lukas[108], Leo-Pekka Lyytikäinen[96,97], Anubha Mahajan[104,109], Pamela R. Matias-Garcia[63,64,110], Christa Meisinger[111,112], Thomas Meitinger[89,113,114], Olle Melander[115], Yuri Milaneschi[116], Pashupati P. Mishra[96,97], Nina Mononen[96,97], Andrew P. Morris[48,104,117], Josyf C. Mychaleckyj[118], Girish N. Nadkarni[32,119], Mariko Naito[71,120], Masahiro Nakatochi[121], Mike A. Nalls[122,123], Matthias Nauck[6,124], Kjell Nikus[125,126], Boting Ning[127], Ilja M. Nolte[128], Teresa Nutile[45], Michelle L. O'Donoghue[129,130], Jeffrey O'Connell[131], Isleifur Olafsson[132], Marju Orho-Melander[133], Afshin Parsa[134,135], Sarah A. Pendergrass[136], Brenda W. J. H. Penninx[116], Mario Pirastu[137], Michael H. Preuss[32], Bruce M. Psaty[138], Laura M. Raffield[139], Olli T. Raitakari[140,141,142], Myriam Rheinberger[2,83,84], Kenneth M. Rice[28], Federica Rizzi[143,144], Alexander R. Rosenkranz[145], Peter Rossing[19,146], Jerome I. Rotter[147], Daniela Ruggiero[45,46], Kathleen A. Ryan[148], Charumathi Sabanayagam[41,42], Erika Salvi[143,149], Helena Schmidt[60], Reinhold Schmidt[74], Markus Scholz[76,77], Ben Schöttker[34,35], Christina-Alexandra Schulz[133], Sanaz Sedaghat[38,150], Christian M. Shaffer[37], Karsten B. Sieber[151], Xueling Sim[13], Mario Sims[152], Harold Snieder[128], Kira J. Stanzick[1], Unnur Thorsteinsdottir[11,153], Hannah Stocker[34,35], Konstantin Strauch[154,155,156], Heather M. Stringham[30], Patrick Sulem[11], Silke Szymczak[58,157], Kent D. Taylor[147], Chris H. L. Thio[128], Johanne Tremblay[68,70,69], Simona Vaccargiu[137], Pim van der Harst[158,159,160], Peter J. van der Most[128], Niek Verweij[158], Uwe Völker[6,161], Kenji Wakai[71], Melanie Waldenberger[63,64,89], Lars Wallentin[162,163], Stefan Wallner[164], Judy Wang[57], Dawn M. Waterworth[151], Harvey D. White[165], Cristen J. Willer[10,166,167], Tien-Yin Wong[41,42], Mark Woodward[40,168,49], Qiong Yang[127], Laura M. Yerges-Armstrong[151], Martina Zimmermann[1], Alan B. Zonderman[56], Tobias Bergler[2], Kari Stefansson[11,153], Carsten A. Böger[2,83,84], Cristian Pattaro[59,171], Anna Köttgen[15,49,171], Florian Kronenberg[169,171], Iris M. Heid[1,171]

[1]Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany. [2]Department of Nephrology, University Hospital Regensburg, Regensburg, Germany. [3]K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology, Trondheim, Norway [4]MRC Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom [5]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [6]DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald, Germany. [7]Department of Population Medicine and Lifestyle Diseases Prevention, Medical University of Bialystok, Bialystok, Poland [8]Department of Clinical and Molecular Medicine, NTNU, Norwegian University of Science and Technology, Trondheim, Norway [9]BioCore - Bioinformatics Core Facility, Norwegian University of Science and Technology, Trondheim. Norway [10]Department of Internal Medicine, Division of Cardiology, University of Michigan, Ann Arbor, MI 48109, USA [11]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. [12]Statistical Consulting Unit StaBLab, Department of Statistics, LMU Munich, Munich, Germany. [13]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore. [14]Department of Public Health Sciences, Loyola University Chicago, Maywood, IL, USA. [15]Institute of Genetic Epidemiology, Department of Biometry, Epidemiology and Medical

Bioinformatics, Faculty of Medicine and Medical Center–University of Freiburg, Freiburg, Germany. [16]Renal Division, Department of Medicine IV, Faculty of Medicine and Medical Center–University of Freiburg, Freiburg, Germany. [17]Memory Impairment and Neurodegenerative Dementia (MIND) Center, University of Mississippi Medical Center, Jackson, MS, USA. [18]Division of Nephrology, Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. [19]Steno Diabetes Center Copenhagen, Gentofte, Denmark. [20]The Bioinformatics Center, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [21]Division of Family Medicine and Primary Care, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden. [22]School of Health and Social Studies, Dalarna University, Stockholm, Sweden. [23]Department of Endocrinology, Clinic of Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway [24]Division of Nephrology, Department of Internal Medicine, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [25]Division of Nephrology, University of Washington, Seattle, WA, USA. [26]Kidney Research Institute, University of Washington, Seattle, WA, USA. [27]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA. [28]Department of Biostatistics, University of Washington, Seattle, WA, USA. [29]Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy. [30]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. [31]Human Genetics Center, University of Texas Health Science Center, Houston, TX, USA. [32]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [33]Digital Health Center, Hasso Plattner Institute and University of Potsdam, Potsdam, Germany. [34]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. [35]Network Aging Research, Heidelberg University, Heidelberg, Germany. [36]Clinic of Thoracic and Occupational Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway [37]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA. [38]Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. [39]Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands. [40]The George Institute for Global Health, University of New South Wales, Sydney, Australia. [41]Singapore Eye Research Institute, Singapore National Eye Center, Singapore, Singapore. [42]Ophthalmology and Visual Sciences Academic Clinical Program (Eye ACP), Duke - NUS Medical School, Singapore, Singapore. [43]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore. [44]Genetics, Merck & Co., Inc, Kenilworth, NJ, USA. [45]Institute of Genetics and Biophysics 'Adriano Buzzati-Traverso'-CNR, Naples, Italy. [46]IRCCS Neuromed, Pozzilli, Italy. [47]Institute for Maternal and Child Health, IRCCS 'Burlo Garofolo', Trieste, Italy. [48]Department of Health Data Science, University of Liverpool, Liverpool, UK. [49]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. [50]Institute of Biomedical Technologies, National Research Council of Italy, Milan, Italy. [51]Bio4Dreams–Business Nursery for Life Sciences, Milan, Italy. [52]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany. [53]Department of Nephrology and Medical Intensive Care, Charité - Universitätsmedizin Berlin, Berlin, Germany [54]Department of Nephrology and Hypertension, Friedrich Alexander University Erlangen-Nürnberg (FAU), Erlangen, Germany. [55]Department of Anatomy and Cell Biology, University Medicine Greifswald, Greifswald, Germany. [56]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, US National Institutes of Health, Baltimore, MD, USA. [57]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA. [58]Institute of Medical Informatics and Statistics, Kiel University, University Hospital Schleswig-Holstein, Kiel, Germany. [59]Eurac Research, Institute for Biomedicine (affiliated with the University of Lübeck), Bolzano, Italy. [60]Institute of Molecular Biology and Biochemistry, Center for Molecular Medicine, Medical University of Graz, Graz, Austria. [61]Department of Genetics, School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran. [62]Molecular Geriatrics, Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden. [63]Research Unit Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [64]Institute of Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany. [65]German Center for Diabetes Research (DZD), Neuherberg, Germany. [66]Iceland School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland [67]Department of Nephrology, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway [68]Montreal University Hospital Research Center, CHUM, Montreal, Quebec, Canada. [69]Medpharmgene, Montreal, Quebec, Canada. [70]CRCHUM, Montreal, Canada. [71]Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan. [72]Kidney Health Research Institute (KHRI), Geisinger, Danville, PA, USA. [73]Department of Nephrology, Geisinger, Danville, PA, USA. [74]Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria. [75]Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria. [76]Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany. [77]LIFE Research Center for Civilization Diseases, University of Leipzig, Leipzig, Germany. [78]Department of Pediatrics, Tampere University Hospital, Tampere, Finland. [79]Department of Pediatrics, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [80]NHLBI's Framingham Heart Study, Framingham, MA, USA. [81]Population Sciences Branch, Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA [82]Geisinger Research, Biomedical and Translational Informatics Institute, Rockville, MD, USA. [83]Department of Nephrology and Rheumatology, Kliniken Südostbayern, Traunstein, Germany. [84]KfH Kidney Center Traunstein [85]Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland. [86]Department of Clinical Physiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [87]Genome Institute of Singapore, Agency for Science Technology and Research, Singapore, Singapore. [88]Deutsches Herzzentrum München, Technische Universität München, Munich, Germany. [89]DZHK (German Center for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich, Germany. [90]Institute of Epidemiology and Medical Biometry, University of Ulm, Ulm, Germany. [91]Division of Nephrology and Hypertension, Loyola University Chicago, Chicago, IL, USA. [92]Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology, Diabetology), Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [93]Department of Medicine, Kuopio University Hospital, Kuopio, Finland. [94]Centre for Medicine and Clinical Research, University of Eastern Finland School of Medicine, Kuopio, Finland. [95]Division of Biomedical Informatics and Personalized Medicine, School of Medicine, University of Colorado Denver-Anschutz Medical Campus, Aurora, CO, USA. [96]Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland. [97]Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [98]Division of Nephrology and Hypertension, Department of Medicine, University of Utah, Salt Lake City, USA. [99]Institute of Epidemiology and Biobank Popgen, Kiel University, Kiel, Germany. [100]A list of members and affiliations appears in the Supplementary Online Material. [101]Department of Medical Sciences, Uppsala University, Uppsala, Sweden. [102]Nuffield Department of Population Health, University of Oxford, Oxford, UK. [103]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [104]Wellcome Center for Human Genetics, University of Oxford, Oxford, UK. [105]Nuffield Dept. of Women's & Reproductive Health, University of Oxford, Level 3, Women's Centre, John Radcliffe Hospital, Oxford, OX3 9DU. [106]Li Ka Shing Centre for Health Information and Discovery, The Big Data Institute, University of Oxford, Oxford, UK [107]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [108]Clinical Sciences, GlaxoSmithKline, Albuquerque, NM, USA. [109]Oxford Center for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. [110]TUM School of Medicine, Technical University of Munich, Munich, Germany. [111]Independent Research Group Clinical Epidemiology, Helmholtz Zentrum München, German Research

2

Center for Environmental Health, Neuherberg, Germany. [112]Chair of Epidemiology, University of Augsburg, University Hospital Augsburg [113]Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany. [114]Institute of Human Genetics, Technische Universität München, Munich, Germany. [115]Hypertension and Cardiovascular Disease, Department of Clincial Sciences Malmö, Lund University, Malmö, Sweden. [116]Department of Psychiatry, Amsterdam Public Health and Amsterdam Neuroscience, Amsterdam UMC/Vrije Universiteit and GGZ inGeest, Amsterdam, the Netherlands. [117]Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, The University of Manchester, Manchester, UK [118]Center for Public Health Genomics, University of Virginia, Charlottesville, Charlottesville, VA, USA. [119]Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [120]Department of Oral Epidemiology, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan [121]Public Health Informatics Unit, Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Nagoya, Japan [122]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. [123]Data Tecnica International, Glen Echo, MD, USA. [124]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. [125]Department of Cardiology, Heart Center, Tampere University Hospital, Tampere, Finland. [126]Department of Cardiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [127]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. [128]Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [129]Cardiovascular Division, Brigham and Women's Hospital, Boston, MA, USA. [130]TIMI Study Group, Boston, MA, USA. [131]University of Maryland School of Medicine, Baltimore, MD, USA. [132]Department of Clinical Biochemistry, Landspitali University Hospital, Reykjavik, Iceland. [133]Diabetes and Cardiovascular Disease–Genetic Epidemiology, Department of Clincial Sciences in Malmö, Lund University, Malmö, Sweden. [134]Division of Kidney, Urologic and Hematologic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA. [135]Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. [136]Geisinger Research, Biomedical and Translational Informatics Institute, Danville, PA, USA. [137]Institute of Genetic and Biomedical Research, National Research Council of Italy, UOS of Sassari, Li Punti, Sassari, Italy. [138]Cardiovascular Health Research Unit, Department of Medicine, Department of Epidemiology, Department of Health Services, University of Washington, Seattle, WA, USA. [139]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. [140]Centre for Population Health Research, University of Turku and Turku University Hospital, Turku, Finland. [141]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. [142]Research Center of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. [143]Department of Health Sciences, University of Milan, Milano, Italy. [144]ePhood Scientific Unit, ePhood SRL, Milano, Italy. [145]Department of Internal Medicine, Division of Nephrology, Medical University Graz, Graz, Austria. [146]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark [147]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA USA. [148]Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD, USA. [149]Neuroalgology Unit, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milan, Italy. [150]Department of Preventive Medicine, Northwestern University, Feinberg School of Medicine, Chicago, Illinois, USA. [151]Human Genetics, GlaxoSmithKline, Collegeville, Pennsylvania, USA. [152]Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA. [153]Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. [154]Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany. [155]Chair of Genetic Epidemiology, IBE, Faculty of Medicine, Ludwig-Maximilians-Universität München, München, Germany. [156]Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), University Medical Center, Johannes Gutenberg University, Mainz, Germany [157]Institute of Medical Biometry and Statistics, University of Lübeck, University Hospital Schleswig-Holstein, Campus Lübeck, Lübeck, Germany [158]Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [159]Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [160]Durrer Center for Cardiovascular Research, The Netherlands Heart Institute, Utrecht, the Netherlands. [161]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [162]Cardiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden. [163]Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden. [164]Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg, Germany. [165]Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland, New Zealand. [166]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA [167]Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA [168]The George Institute for Global Health, University of Oxford, Oxford, UK. [169]Institute of Genetic Epidemiology, Department of Genetics and Pharmacology, Medical University of Innsbruck, Innsbruck, Austria. [170]These authors contributed equally: Mathias Gorski, Humaira Rasheed and Alexander Teumer. [171]These authors jointly supervised this work: Cristian Pattaro, Anna Köttgen, Florian Kronenberg and Iris M. Heid.

Correspondence

Dr. Iris M. Heid and Dr. Mathias Gorski

Department of Genetic Epidemiology

University of Regensburg

Franz-Josef-Strauß-Allee 11

93053 Regensburg

Phone: +49 941 944-5210

Fax: +49 941 944-5212

iris.heid@klinik.uni-regensburg.de and mathias.gorski@klinik.uni-regensburg.de

ABSTRACT

Estimated glomerular filtration rate (eGFR) reflects kidney function. Progressive eGFR-decline can lead to kidney failure, necessitating dialysis or transplantation. Hundreds of loci from genome-wide association studies (GWAS) for eGFR help explain population cross section variability. Since the contribution of these or other loci to eGFR-decline remains largely unknown, we derived GWAS for annual eGFR-decline and meta-analyzed 62 longitudinal studies with eGFR assessed twice over time in all 343,339 individuals and in high-risk groups. We also explored different covariate adjustment. Twelve genome-wide significant independent variants for eGFR-decline unadjusted or adjusted for eGFR-baseline (11 novel, one known for this phenotype), including nine variants robustly associated across models were identified. All loci for eGFR-decline were known for cross-sectional eGFR and thus distinguished a subgroup of eGFR loci. Seven of the nine variants showed variant-by-age interaction on eGFR cross section (further about 350,000 individuals), which linked genetic associations for eGFR-decline with age-dependency of genetic cross-section associations. Clinically important were two to four-fold greater genetic effects on eGFR-decline in high-risk subgroups. Five variants associated also with chronic kidney disease progression mapped to genes with functional *in-silico* evidence (*UMOD, SPATA7, GALNTL5, TPPP*). An unfavorable versus favorable nine-variant genetic profile showed increased risk odds ratios of 1.35 for kidney failure (95% confidence intervals 1.03-1.77) and 1.27 for acute kidney injury (95% confidence intervals 1.08-1.50) in over 2000 cases each, with matched controls). Thus, we provide a large data resource, genetic loci, and prioritized genes for kidney function decline, which help inform drug development pipelines revealing important insights into the age-dependency of kidney function genetics.

KEYWORDS: acute kidney injury, diabetes, chronic kidney disease, gene expression

INTRODUCTION

Glomerular filtration rate (GFR) is accepted as best overall index of kidney function[1]. A GFR<60 mL/min/1.73m^2 defines chronic kidney disease (CKD)[2], which affects about 10% of adults[3]. A decline in GFR over time is characteristic for CKD-progression, which can lead to kidney failure[4] requiring dialysis or kidney transplantation with a high risk of premature mortality[5]. In population studies on kidney function, estimated GFR (eGFR) is usually derived from serum creatinine[6] and annual eGFR-decline as the difference between two such assessments divided by the years between these assessments. Decline in eGFR is age-related, with a physiological loss of ~1 mL/min/1.73m^2 per year[2] generally and 3 mL/min/1.73m^2 per year in the presence of diabetes mellitus (DM), a major risk factor for CKD-progression[7,8]. Therapeutic options to decelerate kidney function decline are limited. In addition to pharmacological inhibitors of the RAAS-system[9], the recent introduction SGLT2 inhibitors show promising reno-protective effects[10,11]. An understanding of the mechanisms of kidney function decline and the developing of new therapeutic options is thus of high clinical and public health relevance[7,12].

Genes underneath genome-wide association study (GWAS) loci for diseases and biomarkers help identify new therapies[13]. Open access GWAS summary statistics from large sample sizes are a highly queried resource, also for causal inference studies[14]. Hundreds of loci and genes are identified by cross-sectional GWAS for eGFR, i.e. GWAS for eGFR based on a single serum creatinine measurement[15–18], which help explain population variability. However, the mechanisms underlying a genetic variant association with lower but stable eGFR over time might not always be disease-relevant. GWAS on parameters more directly linked to disease progression are thought to better inform drug development[19].

Current evidence from GWAS on annual eGFR-decline is limited, owed to substantial logistics in conducting longitudinal studies and thus small sample sizes. Only one variant, in the *UMOD-PDILT* locus, has been identified at genome-wide significance[20] (n~60,000). With an estimated heritability of 38% for annual eGFR-decline[20], comparable to 33%-39% estimated for cross-sectional eGFR in general populations[21,22], much more can be expected in larger sample sizes. Further three loci were genome-wide significant in an extreme phenotype approach, comparing individuals with large eGFR-decline or steep drop into CKD with respective controls[23]. While these are important binary clinical endpoints, methodological literature supports the use of regression methods on undichotomized variables[24].

The limited availability of longitudinal GWAS is not only an issue for kidney function decline, but also generally: e.g. change in lung function (n=27,249[25]), glucose (n=13,807[26]), or blood pressure (n=33,720[27]); consequently, locus findings on biomarker change are few and often unstable[14]. A challenge beyond power is limited experience in longitudinal GWAS with regard to covariate adjustment: clinical trials for disease-related biomarker change require

control for differences in baseline levels between therapy groups[28]. However, covariate adjustment in GWAS requires a careful choice[29]: it can reveal important mediator effects (e.g. DM adjusted for BMI[30]), alter the phenotype (e.g. waist-to-hip ratio "unexpected" by body-mass-index[29,31]), yield artefacts from heritable covariates (collider bias[29]) or non-sense association (e.g. sex adjusted for height[32]). The impact of covariate adjustment on longitudinal GWAS on eGFR-decline, and biomarker change generally, is not well explored.

We thus aimed to identify genetic loci associated with annual eGFR-decline and CKD-progression (defined as eGFR-decline among individuals with CKD at baseline) and to prioritize genes that may inform drug development for slowing down eGFR-decline and CKD-progression. We also aimed to fill the gap of large-data genome-wide SNP summary statistics for annual eGFR-decline and CKD-progression, to help future meta-analyses and Mendelian randomization studies. Finally, we wanted to understand the impact of different covariate adjustment and whether a SNP associated with eGFR-decline showed an age-dependent association on eGFR cross-sectionally (i.e. SNP-by-age interaction on eGFR cross-sectionally). By this, we aimed to contribute to a better understanding of the interpretation of genetic findings for eGFR-decline and other progression traits.

To achieve these aims, we (i) increased sample size for GWAS on annual eGFR-decline to >340,000 individuals based on the CKDGen consortium[33] and UK Biobank[34], (ii) applied a suite of covariate adjustment models, (iii) analyzed SNP-by-age interaction on eGFR cross-sectionally in >350,000 individuals independent of the GWAS on decline, and (v) conducted genetic risk score (GRS) analyses for acute kidney injury (AKI) and end-stage-kidney disease (ESKD).

METHODS

We conducted GWAS meta-analysis based on study-specific summary statistics. Each study utilized data on two measurements of serum creatinine over time and genome-wide SNP-information imputed to 1000 Genomes[35] phase 1 or phase 3, the Haplotype Reference Consortium[36] v1.1 or similar (**Table S1&S2**). Serum creatinine measured at baseline and follow-up were used to estimate eGFR at baseline and follow-up, respectively, according to the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation[6]. Annual eGFR-decline was defined as "-(eGFR at follow-up - eGFR at baseline) / number of years of follow-up". GWAS analyses were conducted separately by ancestry (if applicable), where ancestry was defined by genetic principal components or participants' self-report. GWAS were based on linear regression with different covariate adjustment conducted overall and focused on individuals with DM or CKD at baseline.

Study-specific genome-wide summary statistics and detailed phenotype information were transferred to the meta-analysis center. For each SNP, summary statistics were pooled

and genomic control corrected. Significant genetic variants were identified and respective locus regions selected.

Additionally, we investigated identified SNPs for SNP-by-age interaction on cross-sectional eGFR (based on creatinine or cystatin C, eGFRcrea, eGFRcys) using UK Biobank data that was independent of the SNP identification step (excluding the individuals in the decline GWAS). We computed the GRS and its association on eGFR-decline in the HUNT study via linear regression and provided odds ratios (OR) for GRS association in case-control studies on AKI and ESKD via logistic regression.

Detailed methods are provided in the **Supplementary Methods**.

RESULTS

Overview across studies and models for GWAS

This GWAS meta-analysis included 343,339 individuals from 62 studies (**Supplementary Table S1&S3, Supplementary Figure S1, Methods**) and 12,403,901 analyzable SNPs. Most studies were population-based (76%) and of European ancestry (74%). Study-specific median annual eGFR-decline was independent of sample size and follow-up length (**Supplementary Figure S2A&S2B**) and the median across studies was 1.32 mL/min/1.73m² per year; follow-up length was 1-21 years (median [25th, 75th] = 5 years [4,7]); median age ranged from 33 to 77 years (**Supplementary Figure S2C**).

All analyses were adjusted for age-, sex, and study-specific covariates, which is not mentioned further from here on (stable across different modes of age-adjustment, **Supplementary Figure S3**). We had five GWAS results for eGFR-decline (**Methods**): (i) "unadjusted", (ii) "DM-adjusted", (iii) "adjusted for eGFR-baseline", (iv) restricted to individuals with DM at baseline (unadjusted), and (v) restricted to individuals with CKD at baseline (unadjusted).

Similarities and differences across different model adjustments

There is, to date, no standard conduct for GWAS on eGFR-decline with regard to covariate adjustment. We explored the impact of two potentially important covariates additional to age and sex: (i) DM, as an important risk factors for eGFR-decline and potential mediator, and (ii) eGFR at baseline, as adjustment for baseline levels in analyses of change over time has noted pros (larger effects, better detectability) and cons (biased effects)[37,38].

With regard to DM-adjustment, this model was computed in all studies (n=343,339; 62 studies) and compared to unadjusted results for a subset of studies of varying scope (n=103,970). DM-adjusted SNP-associations on eGFR-decline were precisely the same as unadjusted, in terms of beta-estimates and standard errors (**Supplementary Figure S4A, Supplementary Note S1)**. We therefore did not distinguish these two models further.

7

In contrast, adjustment for eGFR-baseline altered SNP-associations on eGFR-decline (**Supplementary Figure S4B**). Therefore, results from both eGFR-decline unadjusted and adjusted for eGFR-baseline were evaluated in the following. GWAS summary statistics for eGFR-decline adjusted for eGFR-baseline were formula-derived from GWAS summary statistics for unadjusted eGFR-decline and for eGFR-baseline together with study-specific phenotypic information (**Supplementary Note S2**). In a subset of studies (n=103,970), we validated that the formula-approach worked very well in our setting (**Supplementary Note S3**, **Supplementary Figure S4C&D**). Meta-analysis yielded GWAS results for eGFR-decline adjusted for eGFR-baseline for 320,737 individuals (50 studies, **Supplementary Figure S1**).

Twelve variants identified for eGFR-decline unadjusted or adjusted for eGFR-baseline
First, our genome-wide screen for eGFR-decline unadjusted for eGFR-baseline (n=343,339) identified two genome-wide significant independent variants near *UMOD-PDILT* ($P_{DECLINE}$<5x10^{-8}; **Figure 1A**, **Table 1A**): rs34882080, highly correlated with rs12917707 identified previously for this phenotype (r²=1.00)[20], and rs77924615, known for altering *UMOD* expression and urine uromodulin[15] and genome-wide significant for eGFR-decline for the first time.

Second, we evaluated the 263 additional lead variants known for cross-sectional eGFR GWAS[15] for association with baseline-unadjusted eGFR-decline (candidate approach); we had a prior hypothesis that cross-sectionally known variants might also show association with eGFR-decline. We identified two additional variants for eGFR-decline near *PRKAG2* and *SPATA7*, both new loci for this phenotype, at Bonferroni-corrected significance ($P_{DECLINE}$<0.05/263=1.90x10^{-4}; **Table 1A**).

Third, our genome-wide screen for eGFR-decline adjusted for eGFR-baseline (n=320,737) identified 12 independent variants across 11 loci ($P_{DECLINE\_adj\_BL}$<5x10^{-8}, **Figure 1B**), including the four variants already identified by the baseline-unadjusted analyses (directly or via high correlation, r²≥0.9). The 8 variants additionally identified pointed to novel loci for this phenotype. Of these, 5 variants also showed directionally consistent, significant association for eGFR-decline unadjusted for eGFR-baseline (Bonferroni-corrected, $P_{DECLINE}$<0.05/12=4.17x10^{-3}; near *FGF5*, *OVOL1*, *TPPP*, *C15ORF54*, and *ACVR2B*; **Table 1B**), but 3 variants did not ($P_{DECLINE}$ from 0.156 to 0.710; near *GATM*, *CPS1*, *SHROOM3*, **Table 1C**).

Overall, we found 12 variants across 11 loci with genome-wide significant association for eGFR-decline unadjusted and/or adjusted for eGFR-baseline ($P_{DECLINE}$ or $P_{DECLINE\_adj\_BL}$<5x10^{-8}). All but one variant/locus were novel for this phenotype. All resided in loci known for eGFR cross-sectional GWAS[22], but none was associated with DM-status (**Supplementary Table S4**).

The 12 variants' associations showed no between-ancestry heterogeneity, stable statistics in various sensitivity analyses, and no impact by DM-adjustment (**Supplementary Table S5&S6**). Meta-analysis restricted to African American (n=9,038) did not identify associations for published *APOL1* risk variants[39], but two other suggestive variants (**Supplementary Table S7**).

The 12 variants included 9 variants with non-zero effects on eGFR-decline unadjusted for eGFR-baseline (i.e. Bonferroni-corrected significant, i.e. $P_{DECLINE}$<4.17x10^{-3}).

SNP-effects for eGFR-decline were larger when baseline-adjusted than baseline-unadjusted

Several interesting aspects emerged when comparing genetic effect sizes of the 12 identified variants across models. First, we observed consistently larger effects for eGFR-decline baseline-adjusted than baseline-unadjusted (**Figure 2A**), also when restricting to studies where the baseline-adjusted model was directly computed (inserted small panel, **Figure 2A**). This, together with the smaller standard errors (**Supplementary Figure S4B**), explained the larger yield of genome-wide significant loci in the baseline-adjusted GWAS.

Second, we contrasted effect sizes for eGFR-decline unadjusted for eGFR-baseline with those for cross-sectional eGFR[15] (**Figure 2B**). Three variants showed relatively extreme cross-sectional effects and no effect on decline (near *GATM*, *SHROOM3*, *CPS1*). For the other 9 variants, the faster-decline allele was always the cross-sectional eGFR-lowering allele (Spearman correlation coefficient=-0.32). A similar more schematic presentation (**Figure 2C**) illustrates the mathematical relationship between baseline-adjusted and baseline-unadjusted effect sizes (**Supplementary Note S4**). This yields a corollary on the directionality of baseline-adjusted effect sizes: when the faster-decline allele (i.e. $\hat{\beta}_{DECLINE} > 0$) coincides with the baseline eGFR-lowering allele (i.e. $\hat{\beta}_{BL} < 0$), then the baseline-adjusted eGFR-decline effect size is larger than baseline-unadjusted (i.e. $\hat{\beta}_{DECLINE\_adj\_BL} > \hat{\beta}_{DECLINE}$) – in theory. Our data confirmed this empirically (**Figure 2A**). The larger genetic effect sizes for eGFR-decline adjusted for eGFR-baseline are thus a direct consequence of the phenotypic and genetic correlation between eGFR-decline and eGFR-baseline. The genetic effect for eGFR-decline unadjusted for eGFR-baseline provides the relevant effect size for further use and to distinguish between a "genuine association with eGFR-decline" (9 variants) and a pure "collider bias" effect (3 variants).

Four genes with compelling biological in-silico evidence mapped to novel eGFR-decline loci

All 11 identified loci for eGFR-decline coincided with loci detected for cross-sectional eGFR: among the 12 identified variants, 11 variants were genome-wide significant for cross-sectional

eGFR[15] and the variant near *TPPP* showed P=7.63x10^{-6} cross-sectionally with genome-wide significant variants nearby (**Supplementary Figure S5A-C**, **Supplementary Note S5**).

The 8 loci with genuine association for eGFR-decline included the well-known *UMOD-PDILT* locus. Biological evidence at the other seven loci was summarized using the Gene PrioritiSation tool[18] generated from GWAS data on cross-sectional eGFR including evidence for SNP-modulated gene expression (eQTL, false-discovery-rate < 0.05): four lead variants or highly correlated proxies were eQTLs in tubule-interstitial kidney tissue with upregulating effects for *SPATA7* and *GALNTL5* (in *PRKAG2* locus, kidney-tissue specific), a downregulating effect for *FGF5* (kidney-tissue specific), and an upregulating effect for *TPPP* using NEPTUNE[40]. This supported these four genes in novel loci for eGFR-decline as kidney-tissue relevant and potentially causal genes for the association signals.

SNPs for eGFR-decline showed SNP-by-age interaction on cross-sectional eGFR

In the absence of birth cohort effects, we hypothesized that a SNP associated with eGFR-decline might also show an age-dependent association on cross-sectional eGFR, which is SNP-by-age interaction on cross-sectional eGFR. Of note, the age-effect on eGFR should reflect the age-effect on filtration rate, not on creatinine metabolism, within limits of uncertainty of the CKD-EPI formula[6]. To empirically assess this hypothesis, we tested the identified 12 SNPs for SNP-by-age interaction on cross-sectional eGFRcrea or eGFRcys in UK Biobank data, which was independent from and similarly-sized as the decline GWAS (n=351,462 or 351,601 for eGFRcrea or eGFRcys, respectively; **Methods**). For 8 of the 12 SNPs, we found SNP-by-age interaction for eGFRcrea and/or eGFRcys at Bonferroni-corrected significance ($P_{SNPxage}$<0.05/12=4.17x10^{-3}, **Table 2**). Interaction effect sizes were similar between eGFRcrea and eGFRcys (**Figure 3A**), except for the SNP near *GATM*.

The age-dependency of all SNP-effects and main age-effects were approximately linear (**Supplementary Figure S6**, **Supplementary Note S6**). The SNP-by-age interaction effect size can also be interpreted as the genetically modified age-effect on eGFR. This effect was large: e.g., 5 unfavorable alleles decreased eGFRcys by -0.136 mL/min/1.73m^2 per year, which was ~10% of the overall age-effect on eGFRcys (-1.024 mL/min/1.73m^{2}per year, **Supplementary Note S6**). SNP-by-age interaction effects on eGFRcys were highly correlated with SNP-effects on eGFR-decline (both in units of mL/min/1.73m^2 per allele and year: "per year of age-difference between individuals" and "per year of person's aging", respectively; **Figure 3B**).

There was a noteworthy pattern with regard to presence and direction of SNP-by-age interaction: (i) among the 9 variants with genuine association for eGFR-decline, 7 variants showed significant SNP-by-age interaction on cross-sectional eGFRcys (**Table 2A&B**). All interaction effects were negative, i.e. the cross-sectional SNP-effect became larger (in

10

absolute value) with older age. (ii) Among the three SNPs without genuine association for eGFR-decline, two showed no SNP-by-age interaction; the third (near *GATM*) showed SNP-by-age interaction, but only for eGFRcrea and with positive direction ($\hat{\beta}_{SNPxage}$=+0.138, $P_{SNPxage}$=9.71x10^{-5}). Thus, the *GATM* SNP-effect on cross-sectional eGFRcrea gets smaller (in absolute value) by higher age. This might be explained by *GATM* being the rate-limiting enzyme in creatine synthesis in muscle, age-related loss of muscle mass, and thus decreased creatinine production with increasing age - in line with the lack of interaction with eGFRcys, which is unrelated to muscle mass.

A concept of three classes of SNPs for cross-sectional eGFR distinguished by their eGFR-decline association

Our results suggested that SNPs for eGFR-decline were found among SNPs associated with eGFR cross-sectionally. This motivated the idea of, in theory, three classes of SNP-associations on cross-sectional eGFR (intercept) distinguished their eGFR-decline association unadjusted for eGFR-baseline (slope; **Figure 4**): no association with slope (*class I*), association of the eGFR-baseline lowering allele with flatter slope (*class II*), or association of the eGFR-baseline lowering allele with steeper slope (*class III*).

In our data, we found (i) three of the 12 SNPs as *class I*, in line with the lack of SNP-by-age interaction on eGFR cross-sectionally (judged for eGFRcys). (ii) No variant was *class II*, consistent with the lack of positive SNP-by-age interaction on eGFRcys. (iii) The 9 variants with genuine eGFR-decline association were *class III*, and 7 of these showed negative SNP-by-age interaction on eGFR. Thus, our data supported two classes of genetic effects on eGFR: no association with slope or steeper slope for the eGFR-lowering allele.

Larger SNP-effects for eGFR-decline were observed in high-risk subgroups

Individuals with DM and/or CKD (defined as eGFR<60 mL/min/1.73m²) are at higher risk for CKD-progression and kidney failure, prompting us to quantify SNP-effects on eGFR-decline in these high-risk subgroups (meta-analysis for eGFR-decline unadjusted for eGFR-baseline restricted to DM or CKD at baseline, n= 37,375 or 26,653 respectively, **Methods**). For the 9 variants with genuine eGFR-decline association, we found almost all effects to be two- to four-fold larger in DM or in CKD compared to the overall analysis (**Table 3**, average effect size [mL/min/1.73m²/year and allele]: 0.061 in DM, 0.079 in CKD, compared to 0.030 overall).

To get an idea of the magnitude, we scaled the effects to "per 5 unfavorable average alleles" resulting in a decline of 0.305 in DM, 0.395 in CKD, compared to 0.150 mL/min/1.73m²/year overall. This compared well to the 9-variant weighted GRS effect on eGFR-decline per 5 unfavorable average alleles in the HUNT study (n=2,235 with DM, n=502 with CKD, n=46,328 overall; **Methods**): 0.219 in DM, 0.262 in CKD, and 0.102

11

mL/min/1.73m^2/year overall (one-sided P=1.57x10^{-5}, P=0.0193, and P=1.06x10^{-34}, respectively).

The genetic effect sizes were also larger in the two subgroups when viewed relative to the phenotype variance (on the example of HUNT, **Methods**): rs77924615 variant (*UMOD-PDILT* locus) explained 0.38% of the eGFR-decline variance in DM, 0.47% in CKD, and 0.22% overall; the 9-variants jointly explained 1.14%, 1.48%, and 0.51%, respectively. Of note, the explained variance of eGFR-decline overall was comparable to the explained variance of cross-sectional eGFR (rs77924615: 0.21%; 9 variants: 0.62%), but narrow-sense heritability was smaller (**Supplementary Note S7**).

GALNTL5, *SPATA7*, and *TPPP* were identified as candidates for CKD-progression

Variants associated with CKD-progression and mapped genes might help identify drug targets against disease progression[19]. We queried the 9 SNPs with genuine association for eGFR-decline for significant association with CKD-progression, i.e. whether they still showed significant association with eGFR-decline when focusing on individuals with CKD at baseline (judged at P<0.05/9=5.56x10^{-3}, n up to 26,547). We found five such SNPs: (i) two in the *UMOD-PDILT* locus, which confirmed *UMOD* for a role in CKD-progression, (ii) three SNPs in novel loci for eGFR-decline, which mapped to three genes with eQTL in kidney tissue (*GALNTL5* in *PRKAG2* locus, kidney-tissue specific; *SPATA7*, and *TPPP*), making these compelling candidates as CKD-progression genes.

Unfavorable GRS increased the risk for ESKD and AKI

Finally, we wanted to understand the cumulative impact of the 9 genuine eGFR-decline variants for severe clinical endpoints. We thus evaluated the 9-variant weighted GRS in cases-control studies for ESKD and AKI via logistic regression (n$_{cases}$=2,068 and 3,878, n$_{controls}$=4,640 and 11,634, respectively; **Methods**). The GRS effect per 5 unfavorable average alleles showed a significant OR=1.12 for ESKD (95%CI=0.99-1.23; one-sided P=0.033) and OR=1.18 for AKI (95% CI=1.09-1.27; one-sided P<0.0001 **Table 4**). When comparing the individuals with GRS ≥90th versus ≤10th percentile (i.e. ≥14.6 unfavorable alleles versus ≤8.3 in UK Biobank), we found a significant OR=1.35 for ESKD (95%CI=1.03-1.77, one-sided P=0.0157) and OR=1.27 (95%CI=1.08-1.50, one-sided P=0.002, **Table 4**).

DISCUSSION

Here, we provide data and results on a large longitudinal GWAS on annual eGFR-decline with >340.000 individuals from mostly population-based studies – to our knowledge the largest GWAS on annual eGFR-decline so far and probably one of the largest longitudinal GWAS of any trait. We identified 12 variants across 11 loci as genome-wide significant for annual eGFR-

decline unadjusted and/or adjusted for eGFR-baseline **(Figure 5)**. These included 9 variants across 8 loci with non-zero association unadjusted for eGFR-baseline, which we termed "genuinely" associated with eGFR-decline. Seven of these 9 variants also showed SNP-by-age interaction on cross-sectional eGFR in independent data of >350,000 individuals, while the three variants without genuine association did not. We generated and provide genome-wide summary statistics for eGFR-decline, CKD-progression, and eGFR-decline in DM. This data resource is informative for future meta-analyses, causal inference studies via Mendelian Randomization[41], and drug development pipelines.

Clinically very important is our finding of the two-to four-fold larger genetic effects of almost all identified variants when focusing on individuals with DM or CKD at baseline, since these individuals are already at higher risk of kidney failure. This observation is in line with a "horse-racing effect"[42] ("a faster horse is more likely observed up front"): individuals with an accumulation of faster eGFR-decline alleles are more likely observed with low eGFR at a given point in time, implying that these genetic effects might partly explain lower eGFR at baseline. A part of the larger eGFR-decline effect among CKD individuals might reflect collider bias. However, DM-status does not fulfill the characteristics of a collider for the SNP-associations with eGFR-decline (no impact by adjusting for DM-status, no SNP-association with DM-status), rendering the higher eGFR-decline effects in DM genuine.

The clinical relevance is further underscored by the 9-variant GRS being associated with increased risk of AKI and ESKD. This observation requires further analyses in future larger data. If substantiated, this may indicate a genetic risk of incomplete kidney function recovery after AKI and a genetic predisposition for ESKD.

The 9 identified variants across 8 loci included the *UMOD-PDILT* locus associated with eGFR-decline and CKD-progression, which is largely confirmatory but serves as proof-of-concept. A variant near *MIR378C* previously identified for CKD-progression[43] (n~3000) was not confirmed here. Our other 7 loci are novel for eGFR-decline (near/in *PRKAG2-GALNTL5*, *SPATA7*, *FGF5*, *OVOL1*, *TPPP*, *C15ORF54*, and *ACVR2B*). These included at least three loci associated with CKD-progression (defined as eGFR-decline in individuals with CKD at baseline), mapping to the genes *GALNTL5*, *SPATA7* and *TPPP* by SNP-modulated expression in tubolo-interstitium[15,18]. These associations and genes for CKD-progression are in strong demand as genetic information on a disease progression phenotype, in order to help identify treatment[19]. Our data particularly flags *TPPP* by its locus' large effect on eGFR-decline and CKD-progression, making it second only after *UMOD*. This also documents the value of longitudinal GWAS in revealing relevance of genes like *TPPP*: the *TPPP* locus was one of hundreds of small effect loci cross-sectionally, but among the few loci longitudinally.

Our results highlight some overlap of quantitative eGFR-decline genetics with binary extreme decline genetics[23], but also distinction. All loci identified here were directionally

consistent, nominally significant with "rapid3" and/or "CKDi25" (one-sided P<0.05) and two were genome-wide significant for rapid3 or CKDi25 (*UMOD-PDILT*, *PRKAG2-GALNTL5*). Particularly the loci identified here for CKD-progression, which is among individuals with CKD at baseline, complement the previously reported associations with CKDi25, which is among individuals without CKD at baseline. Methodologically, regression applied to a quantitative rather than dichotomized outcome has larger power and statistical advantages.

While all variants identified for eGFR-decline captured loci known from cross-sectional eGFR[15], these associations are important on various accounts. First, the mere fact that eGFR-decline genetics is a subgroup of cross-sectional eGFR genetics is informative for future searches. Second, the finding that the full genetic signals were the same enabled the use of fine-mapping results from cross-sectional GWAS in >1 million individuals[18] to prioritize genes also for longitudinal eGFR-decline. Third, all faster-decline alleles were the cross-sectional eGFR-lowering alleles. Together, this supported two classes of genetic variants for cross-sectional eGFR, distinguished by lack or presence of a slope effect, with steeper slope for the cross-sectional eGFR-lowering allele. The data rendered the third theoretical option, i.e. presence of a slope effect with flatter slope for the cross-sectional eGFR-lowering allele, void.

Some limitations warrant mentioning. Although this GWAS is currently the largest GWAS on eGFR-decline so far, more loci for eGFR-decline and CKD-progression might be detectable upon further increased sample size. The yield of eGFR-decline loci in >340,000 individuals was comparably low considering older GWAS for cross-sectional eGFR having already detected >50 loci in 170,000 individuals[44]. We used the CKD-EPI formula containing an ancestry term (Levey et al., Ann Intern Med), accounted for by ancestry-specific GWAS; future work should utilize the new ancestry-term-free CKD-EPI formula 2021 (Inker et al., NEJM). Evaluating the potential existence of sex-specific genetic effects on eGFR-decline is of interest, but was not addressed in this project. The target population is primarily population-based, including kidney diseases proportional to respective prevalence, and primarily European ancestry. Larger all-ancestry meta-analyses on eGFR-decline will open up opportunities to also utilize differential linkage disequilibrium between ancestries to help narrow down causal variants and genes. The interpretability of the SNP-by-age interaction on cross-sectional eGFR is limited to the age spectrum in the data (40-70 years) and by the power given the sample size; still, the sample size used was large and the age range typical also for most eGFR-decline GWAS studies. Two aspects need mentioning regarding the phenotype definition: uncertainty in eGFR-decline may be larger for studies with shorter follow-up, which decreases power, but measurement error in the outcome does not induce bias in linear regression[45]. By defining annual eGFR-decline from two eGFR assessments over time, our SNP associations capture only the linear component of decline. Serial eGFR assessments are better to characterize eGFR-trajectories, but at the cost of limiting sample size, since such

studies are few and typically small. Furthermore, generalized additive mixed models for non-linear eGFR-trajectories are complex and require particularly large sample sizes. The linear modelling of eGFR-decline is a reasonable approximation of monotonous decline, maintaining large sample sizes and limiting model complexity to be applicable for GWAS. Overall, the choice of the adjustment, target population, and phenotype definition are important to consider when interpreting results. While some modelling aspects are addressed here, other covariate adjustment or relative decline as phenotype might reveal further or other genetic loci. Future work is warranted to quantify effects in different target populations and the genetically determined shape of the decline, which requires more – and larger – longitudinal studies, ideally with more than two eGFR assessments over time.

Methodologically unique is our contrasting of GWAS SNP-associations on eGFR-decline for different covariate adjustment, which fills an important gap and helps design future studies. This is highly relevant, since covariate adjustment can alter GWAS findings and interpretation[29–32,46]. Adjusting for baseline DM-status had no impact, but genetic effects for eGFR-decline were larger when restricting to DM-individuals; this suggests DM-status as modulator for the SNP-association with eGFR-decline rather than mediator (i.e. in the causal pathway from SNP to eGFR-decline) or collider (i.e. generating biased association). Adjustment for eGFR-baseline yielded larger eGFR-decline effects and more genome-wide significant variants. Glymour et al. highlight that adjustment for baseline levels in analyses of change may help detect effects, but can induce spurious associations when the rate of change observed after baseline reflects a rate of change experienced in the past[37]. This might reflect the situation here rendering the larger genetic effects adjusted for eGFR-baseline - and the larger genetic effects when restricting to individuals with CKD at baseline – reflective of collider bias. Glymour et al. recommend the documentation of change effects without baseline adjustment[37]. In line with this, we considered a variant's association with eGFR-decline genuine, when the variant reached genome-wide significance baseline-unadjusted or baseline-adjusted and Bonferroni-corrected significance baseline-unadjusted. The baseline-unadjusted model provides the relevant genetic effect sizes for eGFR-decline.

Interestingly, two of the three associations without genuine eGFR-decline association may relate to biomarker generation rather than kidney function: *GATM* and *CPS1*, known for a role in creatine biosynthesis[42] and urea cycle[43], respectively, reside in loci without supporting association with cross-sectional cystatin-based eGFR[18]. Conversely, the *SHROOM3* locus was associated with cystatin-based eGFR[18,22] and experimental studies support a role of *SHROOM3* in kidney pathology[47–49]; thus, *SHROOM3* appears to have an effect on cross-sectional kidney function, but not on kidney function decline within the limits of detectability by sample size.

A further unique aspect of our work is the empirical evidence for a link between SNP-effects on eGFR-decline with SNP-by-age interaction effects on cross-sectional eGFR. By this, we provide important insights into the age-dependency of kidney function genetics as well as into the genetic dependency of aging eGFR in adult general populations, where "aging" includes onset of age-related diseases as they develop in populations. Considering the much broader availability of cross-sectional than longitudinal data, the further parallel exploitation of SNP-by-age interaction might be a promising route to help improve our understanding of the mechanisms of kidney function decline over time.

In summary, we provide GWAS summary statistics, identified genetic loci, and prioritized genes for kidney function decline and CKD-progression. While *UMOD* has drawn attention already, *GALNTL5*, *SPATA7*, and *TPPP* may now receive more focus as therapeutic targets for disease progression. Our exploration of different covariate adjustment and the comparison to age-dependency of SNP-effect on eGFR cross-sectional provides important insights into the interpretation of these effects. With the emerging large biobank data linking medical records, longitudinal GWAS will become very important in the future. Our methodological framework is informative and applicable also generally for longitudinal phenotypes.

Availability of data and materials

To support future work, we provide genome-wide summary statistics on eGFR-decline unadjusted for eGFR-baseline (adjusted for age, sex and DM-status) overall and restricted to individuals with DM or CKD at baseline (all adjusted for age and sex) (https://www.uni-regensburg.de/decline and http://ckdgen.imbi.uni-freiburg.de). The summary statistics on eGFR-decline in individuals with CKD at baseline can be considered genetic effects on CKD-progression. We also provide genome-wide summary statistics on eGFR-decline adjusted for eGFR-baseline (additionally to adjustment for age and sex), but these summary statistics should be used with great care and an understanding that beta-estimates are subject to collider bias. For quantification of the genetic effect on eGFR-decline, the results unadjusted for eGFR-baseline should be utilized.

DISCLOSURE STATEMENT

SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

Supplementary Methods

Note S1. Equivalence of DM-adjusted versus not DM-adjusted GWAS on eGFR-decline in the validation meta-analysis

Note S2. Formula-based covariate adjustment using GWAS summary statistics

Note S3. Validation of the formula-derived association for eGFR-decline adjusted for eGFR-baseline

Note S4. Graphical illustration of the relationship between SNP-effects on eGFR-decline unadjusted and adjusted for eGFR-baseline

Note S5. Comparison of the signals for eGFR-decline unadjusted and adjusted for eGFR-baseline and cross-sectional eGFR for the 11 identified loci

Note S6. Age-dependency of SNP-effects and main age effect on eGFR

Note S7. Narrow-sense heritability

Figure S1. Meta-analysis workflow

Figure S2. Study-specific median annual eGFR-decline versus sample size, follow-up time and median age

Figure S3. Influence of alternative adjustments for age on eGFR-decline in UK Biobank

Figure S4A. No influence from adjusting SNP-associations for eGFR-decline for diabetes mellitus (DM)

Figure S4B. Differences between SNP-association for eGFR-decline unadjusted versus adjusted for eGFR-baseline

Figure S4C. Validation of formula-derived adjustment for eGFR-baseline in eGFR-decline associations (part 1).

Figure S4D. Validation of formula-derived adjustment for eGFR-baseline in eGFR-decline associations (part 2)

Figure S5. Region plots of loci identified for eGFR-decline unadjusted and adjusted for eGFR-baseline

Figure S6. Age-dependency of eGFR and age-dependency of the variant effects on eGFR in UK Biobank

Table S1. Description of participating studies: study design

Table S2. Description of participating studies: genotyping and imputation

Table S3. Description of participating studies: phenotype distribution

Table S4. The 12 identified variants for eGFR-decline were associated with other kidney phenotypes, but not with DM-status

Table S5. The 12 identified variants for eGFR-decline do not show heterogeneity between ancestries and FHS is not an influential study

Table S6. No influence by DM-adjustment versus no DM-adjustment or by model-based versus formula-based adjusting for baseline eGFR (BL) on the 12 variants' association with eGFR-decline

Table S7. Association of *APOL1* risk variants in African American and European CKDGen studies

Extended acknowledgements, study funding information and author contributions

Supplementary References

Author contributions

18

REFERENCES

1. Levey AS, Coresh J, Tighiouart H, Greene T, Inker LA. Measured and estimated glomerular filtration rate: current status and future directions. *Nat Rev Nephrol.* 2020;16(1):51-64.

2. Andrassy KM. Comments on "KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease." *Kidney Int.* 2013;84(3):622-623.

3. Eckardt K-U, Coresh J, Devuyst O, et al. Evolving importance of kidney disease: from subspecialty to global health burden. *Lancet (London, England).* 2013;382(9887):158-169.

4. Neuen BL, Weldegiorgis M, Herrington WG, Ohkuma T, Smith M, Woodward M. Changes in GFR and Albuminuria in Routine Clinical Practice and the Risk of Kidney Disease Progression. *Am J Kidney Dis.* 2021;78(3):350-360.e1.

5. Meguid El Nahas A, Bello AK. Chronic kidney disease: the global challenge. *Lancet (London, England).* 2005;365(9456):331-340.

6. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med.* 2009;150(9):604-612.

7. Levin A, Stevens PE, Bilous RW, et al. Notice. *Kidney Int Suppl.* 2013;3(1):1.

8. Hemmelgarn BR, Zhang J, Manns BJ, et al. Progression of kidney dysfunction in the community-dwelling elderly. *Kidney Int.* 2006;69(12):2155-2161.

9. Garcia Sanchez JJ, Thompson J, Scott DA, et al. Treatments for Chronic Kidney Disease: A Systematic Literature Review of Randomized Controlled Trials. *Adv Ther.* 2022;39(1):193-220.

10. Heerspink HJL, Stefánsson B V., Correa-Rotter R, et al. Dapagliflozin in Patients with Chronic Kidney Disease. *N Engl J Med.* 2020;383(15):1436-1446.

11. Borges-Júnior FA, Silva dos Santos D, Benetti A, et al. Empagliflozin Inhibits Proximal Tubule NHE3 Activity, Preserves GFR, and Restores Euvolemia in Nondiabetic Rats with Induced Heart Failure. *J Am Soc Nephrol.* 2021;32(7):1616-1629.

12. Abbafati C, Abbas KM, Abbasi-Kangevari M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet.* 2020;396(10258):1204-1222.

13. King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet.* 2019;15(12):e1008489.

14. Buniello A, Macarthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47(D1):D1005-D1012.

15. Wuttke M, Li Y, Li M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet.* 2019;51(6):957-972.

16. Hellwege JN, Velez Edwards DR, Giri A, et al. Mapping eGFR loci to the renal transcriptome and phenome in the VA Million Veteran Program. *Nat Commun.* 2019;10(1):3842.

17. Chambers JC, Zhang W, Lord GM, et al. Genetic loci influencing kidney function and chronic kidney disease. *Nat Genet.* 2010;42(5):373-375.

18. Stanzick KJ, Li Y, Schlosser P, et al. Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. *Nat Commun.* 2021;12(1):4350.

19. Paternoster L, Tilling K, Davey Smith G. Genetic epidemiology and Mendelian randomization for informing disease therapeutics: Conceptual and methodological challenges. *PLOS Genet.* 2017;13(10):e1006944.

20. Gorski M, Tin A, Garnaas M, et al. Genome-wide association study of kidney function decline in individuals of European descent. *Kidney Int.* 2015;87(5):1017-1029.

21. Fox CS, Yang Q, Cupples LA, et al. Genomewide linkage analysis to serum creatinine, GFR, and creatinine clearance in a community-based population: the Framingham Heart Study. *J Am Soc Nephrol.* 2004;15(9):2457-2461.

22. Wuttke M, Li Y, Li M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet.* 2019;51(6):957-972.

23. Gorski M, Jung B, Li Y, et al. Meta-analysis uncovers genome-wide significant variants for rapid kidney function decline. *Kidney Int.* 2021;99(4):926-939.

24. MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. *Psychol Methods.* 2002;7(1).

25. Tang W, Kowgier M, Loth DW, et al. Large-scale genome-wide association studies and meta-analyses of longitudinal change in adult lung function. *PLoS One.* 2014;9(7):e100776.

26. Liu C-T, Merino J, Rybin D, et al. Genome-wide Association Study of Change in Fasting Glucose over time in 13,807 non-diabetic European Ancestry Individuals. *Sci Rep.* 2019;9(1):9439.

27. Gouveia MH, Bentley AR, Leonard H, et al. Trans-ethnic meta-analysis identifies new loci associated with longitudinal blood pressure traits. *Sci Rep.* 2021;11(1):4075.

28. Vickers AJ, Altman DG. Statistics notes: Analysing controlled trials with baseline and follow up measurements. *BMJ.* 2001;323(7321):1123-1124.

29. Aschard H, Vilhjálmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am J Hum Genet.* 2015;96(2):329-339.

30. Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science.* 2007;316(5826):889-894.

31. Winkler TW, Günther F, Höllerer S, et al. A joint view on genetic variants for adiposity differentiates subtypes with distinct metabolic implications. *Nat Commun.* 2018;9(1):1946.

32. Day FR, Loh P-R, Scott RA, Ong KK, Perry JRB. A Robust Example of Collider Bias in a Genetic Association Study. *Am J Hum Genet.* 2016;98(2):392-393.

33. Köttgen A, Pattaro C. The CKDGen Consortium: ten years of insights into the genetic basis of kidney function. *Kidney Int.* 2020;97(2):236-242.

34. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12(3):e1001779.

35. The 1000 Genomes Project Consortium. An integrated map of genetic variation Supplementary Material. *Nature.* 2012;135:1-113.

36. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016;48(10):1279-1283.

37. Glymour MM, Weuve J, Berkman LF, Kawachi I, Robins JM. When is baseline adjustment useful in analyses of change? An example with education and cognitive change. *Am J Epidemiol.* 2005;162(3):267-278.

38. Yanez ND, Kronmal RA, Shemanski LR. The effects of measurement error in response variables and tests of association of explanatory variables in change models. *Stat Med*. 1998;17(22):2597-2606.

39. Parsa A, Kao WHL, Xie D, et al. *APOL1* Risk Variants, Race, and Progression of Chronic Kidney Disease. *N Engl J Med*. 2013;369(23):2183-2196.

40. Gillies CE, Putler R, Menon R, et al. An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. *Am J Hum Genet*. 2018;103(2):232-244.

41. Davey Smith G, Paternoster L, Relton C. When Will Mendelian Randomization Become Relevant for Clinical Practice and Public Health? *JAMA*. 2017;317(6):589-591.

42. Peto R. The horse-racing effect. *Lancet (London, England)*. 1981;2(8244):467-468.

43. Parsa A, Kanetsky PA, Xiao R, et al. Genome-wide association of CKD progression: The chronic renal insufficiency cohort study. *J Am Soc Nephrol*. 2017;28(3):923-934.

44. Pattaro C, Köttgen A, Teumer A, et al. Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet*. 2012;8(3):e1002584.

45. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement Error in Nonlinear Models: A Modern Perspective*. Vol 2. 2nd ed.; 2006.

46. Vansteelandt S, Goetgeluk S, Lutz S, et al. On the adjustment for covariates in genetic association analysis: A novel, simple principle to infer direct causal effects. *Genet Epidemiol*. 2009;33(5):394-405.

47. Yeo NC, O'Meara CC, Bonomo JA, et al. Shroom3 contributes to the maintenance of the glomerular filtration barrier integrity. *Genome Res*. 2015;25(1):57-65.

48. Khalili H, Sull A, Sarin S, et al. Developmental Origins for Kidney Disease Due to Shroom3 Deficiency. *J Am Soc Nephrol*. 2016;27(10):2965-2973.

49. Matsuura R, Hiraishi A, Holzman LB, et al. SHROOM3, the gene associated with chronic kidney disease, affects the podocyte structure. *Sci Rep*. 2020;10(1):21103.

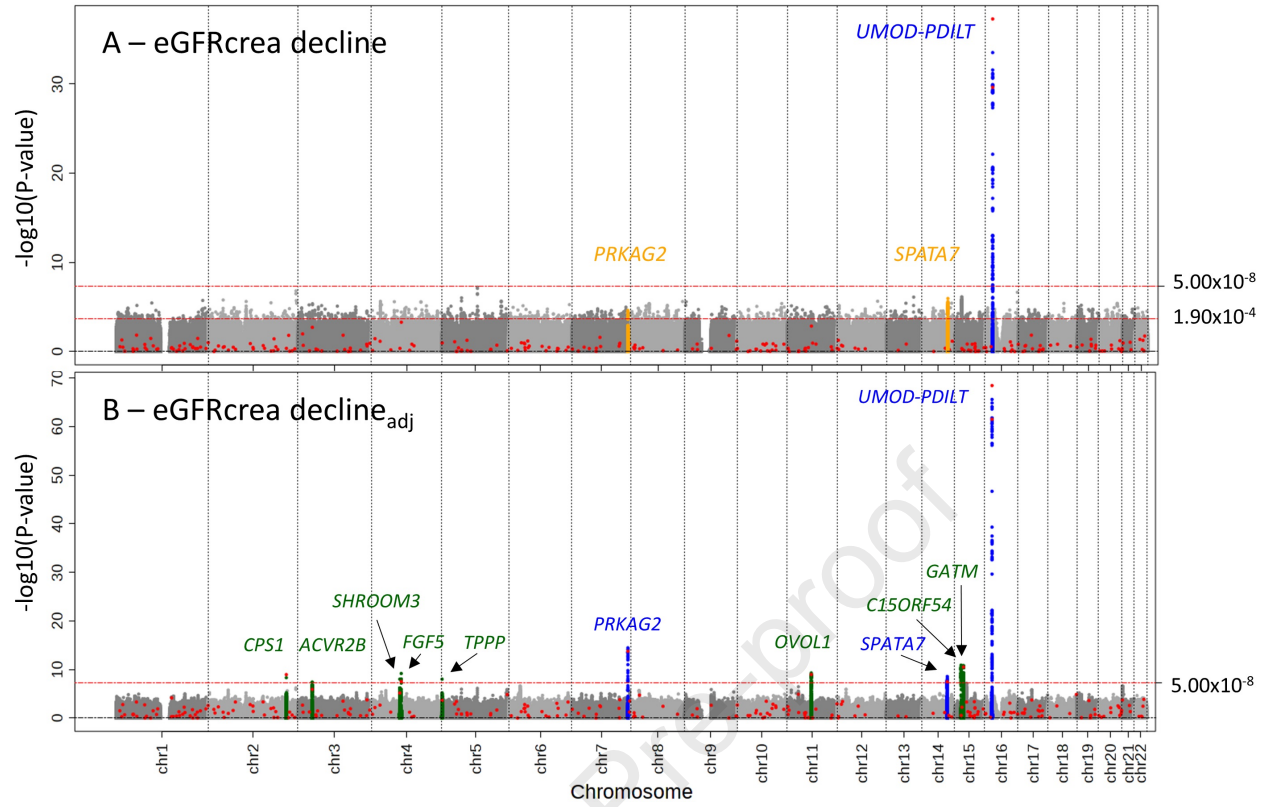**ACKNOWLEDGEMENTS**

Table 1: **Twelve independent variants in 11 loci identified for association with eGFR-decline unadjusted and adjusted for eGFR-baseline.** We conducted GWAS for eGFR-decline baseline-unadjusted and baseline-adjusted ("decline", n up to 343,339; decline$_{adj}$, n up to 320,737). This identified **(A)** 2 variants with genome-wide significance for eGFR-decline baseline-unadjusted (*UMOD-PDILT*, P$_{decline}$<5x10^{-8}) and 2 further variants in a candidate search of the 263 variants known for cross-sectional eGFR[15] outside *UMOD-PDILT*, judged at Bonferroni-corrected significance (P$_{decline}$<0.05/263=1.90x10^{-4}; *PRKAG2*, *SPATA7*), **(B)** 5 variants with genome-wide significance for eGFR-decline baseline-adjusted AND Bonferroni-corrected significant baseline-unadjusted (P$_{decline-adj-BL}$<5x10^{-8}, P$_{decline}$ <0.05/12=4.17x10^{-3}), **(C)** 3 variants with genome-wide significance for eGFR-decline baseline-adjusted but not significantly associated baseline-unadjusted (P$_{decline-adj-BL}$<5x10^{-8}, P$_{decline}$≥4.17x10^{-3}). For each identified variant, we show results for decline (baseline-unadjusted), for decline baseline-adjusted, and for cross-sectional eGFR[15]. Beta-estimates are in mL/min/1.73² per year and per faster-decline allele; significant P-values are stated in bold.

SNPID	Locus Name	Chr	Pos	EA/OA	EAF	decline Beta	decline P	decline$_{adj}$ Beta	decline$_{adj}$ P	cross-sectional Beta	cross-sectional P
A from GWAS/candidate search for decline (baseline-unadjusted)											
rs34882080	*UMOD-PDILT*	16	20,361,441	a/g	0.815	0.065	**2.45x10^{-30}**	0.092	**3.31x10^{-62}**	-0.009	**2.86x10^{-95}**
rs77924615	*UMOD-PDILT*	16	20,392,332	g/a	0.798	0.074	**5.30x10^{-38}**	0.099	**3.75x10^{-69}**	-0.010	**1.45x10^{-138}**
rs10254101	*PRKAG2**	7	151,415,536	t/c	0.276	0.020	**4.10x10^{-05}**	0.037	**1.78x10^{-14}**	-0.007	**1.85x10^{-67}**
rs1028455	*SPATA7**	14	88,829,975	t/a	0.657	0.021	**5.90x10^{-06}**	0.024	**3.43x10^{-08}**	-0.002	**4.78x10^{-10}**
B from GWAS for decline$_{adj}$, with association for decline (baseline-unadjusted)											
rs1458038	*FGF5*	4	81,164,723	c/t	0.690	0.019	**3.87x10^{-05}**	0.028	**6.85x10^{-10}**	-0.003	**7.49x10^{-24}**
rs4930319	*OVOL1*	11	65,555,458	c/g	0.333	0.015	**9.93x10^{-04}**	0.028	**5.27x10^{-10}**	-0.003	**2.21x10^{-24}**
rs434215	*TPPP§*	5	699,046	a/g	0.277	0.020	**3.70x10^{-04}**	0.032	**7.19x10^{-09}**	-0.003	7.63x10^{-06}
rs28857283	*C15ORF54†*	15	39,224,711	g/a	0.656	0.021	**1.47x10^{-06}**	0.030	**1.31x10^{-11}**	-0.002	**6.20x10^{-09}**
rs13095391	*ACVR2B*	3	38,447,232	a/c	0.502	0.017	**1.77x10^{-04}**	0.025	**4.03x10^{-08}**	-0.003	**6.57x10^{-15}**
C from GWAS for decline$_{adj}$, without association for decline (baseline-unadjusted)											
rs9998485	*SHROOM3*	4	77,362,445	a/g	0.466	0.007	0.156	0.027	**9.84x10^{-09}**	-0.005	**1.22x10^{-41}**
rs1047891	*CPS1*	2	211,540,507	a/c	0.293	0.004	0.441	0.029	**1.15x10^{-09}**	-0.007	**1.18x10^{-75}**
rs2453533	*GATM*	15	45,641,225	a/c	0.422	0.002	0.710	0.029	**1.72x10^{-11}**	-0.009	**4.57x10^{-141}**

SNPID=Variant identifier on GRCh37, **Locus name**=Nearest Gene, **Chr** and **Position**=Chromosome and Position on GRCh37, **EA/OA**=Effect allele / other allele, **EAF**=effect allele frequency, **beta** and **P**=genetic effect coefficient of association and association P-value. * In *PRKAG2* and *SPATA7* loci, variants with smallest P$_{decline}$ (rs73158188 and rs7160717, respectively) were highly correlated with these candidate-based variants (r²=1.00 and 0.93, respectively). § Since the *TPPP* locus lead variant had imputation quality <0.6 in 45% of the studies (median 0.64), we analyzed this locus omitting the imputation quality filter (with filter: decline$_{adj}$ beta=0.033, P=1.00x10^{-8}; decline beta=0.015, P=0.039; median imputation quality=0.74). † In the *C15ORF54* locus, the identified lead variant for decline was highly correlated with a 2nd signal lead variant for cross-sectional eGFR (rs28833881, r²=0.90), but not with the 1st signal lead variant (rs12913015, r²=0.04).

Table 2: SNP-by-age interaction for cross-sectional eGFR for the 12 identified variants.
For the 12 identified variants, we conducted SNP-by-age interaction analysis for cross-sectional eGFRcrea and eGFRcys in UK Biobank (excluding individuals from decline GWAS; n=351,462 for eGFRcrea, n=351,601 for eGFRcys; main age effect modelled non-linearly, main SNP effect linearly, age centered at 50 years). The interaction term (age effect and SNP effect modelled linearly) was judged at Bonferroni-corrected significance level ($P<0.05/12=4.17\times10^{-3}$). Beta-estimates are in mL/min/1.73² per year and per cross-sectional eGFR-lowering allele (which was equivalent to faster-decline allele for each SNP); significant P-values are stated in bold.

SNPID	Locus Name	EA/OA	SNP x age interaction eGFRcrea		SNP x age interaction eGFRcys	
			Beta	P	Beta	P
A from GWAS/candidate search for decline (baseline-unadjusted)						
rs34882080	*UMOD-PDILT*	a/g	-0.043	**5.53×10^{-22}**	-0.045	**2.37×10^{-17}**
rs77924615	*UMOD-PDILT*	g/a	-0.050	**2.55×10^{-29}**	-0.054	**6.59×10^{-25}**
rs10254101	*PRKAG2*	t/c	-0.009	0.0263	-0.015	**9.84×10^{-04}**
rs1028455	*SPATA7*	t/a	-0.014	**2.19×10^{-04}**	-0.014	**1.06×10^{-03}**
B from GWAS for decline_{adj}, with association for decline (baseline-unadjusted)						
rs1458038	*FGF5*	c/t	-0.013	**7.11×10^{-04}**	-0.013	**3.12×10^{-03}**
rs4930319	*OVOL1*	c/g	-0.015	**2.55×10^{-05}**	-0.016	**1.84×10^{-04}**
rs434215	*TPPP*	a/g	-0.028	**1.02×10^{-10}**	-0.033	**5.02×10^{-11}**
rs28857283	*C15ORF54*	g/a	-0.010	5.09×10^{-03}	-0.006	0.148
rs13095391	*ACVR2B*	a/c	0.004	0.227	0.002	0.695
C from GWAS for decline_{adj}, without association for decline (baseline-unadjusted)						
rs9998485	*SHROOM3*	a/g	-0.004	0.206	-0.009	0.022
rs1047891	*CPS1*	a/c	0.004	0.228	0.005	0.244
rs2453533	*GATM*	a/c	0.014	**9.71×10^{-05}**	0.002	0.722

SNPID=Variant identifier on GRCh37, **Locus name**=Nearest Gene, **EA/OA**=Effect allele / other allele, **Beta** and **P**=genetic effect and association P-value. The *TPPP* variant rs434215 is well-imputed in the UK Biobank (imputation quality=0.82).

Table 3: The 9 variants' effects on eGFR-decline unadjusted for eGFR-baseline in high-risk subgroups. Shown are the 9 variants with genuine association for eGFR-decline for their association with eGFR-decline restricted to individuals with baseline diabetes mellitus (DM, n up to 38,206) or baseline CKD (i.e. eGFR<60 mL/min/1.73m², n up to 26,653). Beta-estimates and 95% confidence intervals (CI) are in mL/min/1.73m² per year and per faster-decline allele.

SNPID	Locus Name	Decline among DM at baseline		Decline among CKD at baseline		Decline among all	
		Beta	95% CI	Beta	95% CI	Beta	95% CI
A from GWAS/candidate search for decline (baseline-unadjusted)							
rs34882080	*UMOD-PDILT*	0.159*	0.108, 0.211	0.138*	0.074, 0.203	0.065	0.054, 0.076
rs77924615	*UMOD-PDILT*	0.136*	0.084, 0.189	0.167*	0.099, 0.235	0.074	0.063, 0.085
rs10254101	*PRKAG2*	0.065	0.020, 0.110	0.095*	0.042, 0.148	0.020	0.010, 0.030
rs1028455	*SPATA7*	0.030	-0.011, 0.071	0.085*	0.034, 0.135	0.021	0.012, 0.029
B from GWAS for decline$_{adj}$, with association for decline (baseline-unadjusted)							
rs1458038	*FGF5*	0.030	-0.013, 0.072	0.040	-0.013, 0.092	0.019	0.010, 0.028
rs4930319	*OVOL1*	0.021	-0.021, 0.062	0.031	-0.019, 0.080	0.015	0.006, 0.024
rs434215	*TPPP*§	0.031	-0.024, 0.086	0.112*	0.043, 0.180	0.020	0.006, 0.035
rs28857283	*C15ORF54*	0.046	0.005, 0.086	0.042	-0.007, 0.091	0.021	0.013, 0.030
rs13095391	*ACVR2B*	0.029	-0.021, 0.080	0.006	-0.054, 0.066	0.017	0.008, 0.026
Average		**0.061**		**0.079**		**0.030**	

SNPID=Variant identifier on GRCh37, **Locus name**=Nearest Gene, **Beta**=genetic effect of genetic association where the effect alleles is the same as in **Table 1** and **Table 2**, **95% CI** = 95% confidence interval of Beta (Beta±1.96*standard error of the association).

* Statistically significant different from zero (P< 0.05/9=5.56x10^{-3}).

§ Since the lead variant had imputation quality <0.6 in 45% of the studies (median 0.64), we analyzed this variant omitting the imputation quality filter (with filter: decline among DM at baseline beta=-0.093, P=0.338, n=927; decline among eGFR <60 mL/min/1.73m² beta=0.022, P=0.618, n=2924; median imputation quality=0.74).

Table 4: Genetic risk score (GRS) analyses for end-stage kidney disease (ESKD) and Acute Kidney Injury (AKI). In 3 case-control studies for ESKD and one for AKI, we computed the weighted GRS across the 9 eGFR-decline variants (counting the faster-decline alleles, weighted by effect size for eGFR-decline unadjusted for eGFR-baseline; divided by sum of weights and multiplied by 9, i.e. scaled as 0 to 18). Shown are odds ratios (OR), 95% confidence intervals (CI) and P-values (one-sided) for the quantitative GRS association (per 5 "average" unfavorable alleles) and for a high versus low GRS association (≥95th versus ≤5th, ≥90th versus ≤10th GRS percentiles derived in UK Biobank) with (A) ESKD and (B) AKI. Associations are derived by logistic regression adjusted for matching variables age-group and sex (AKI additionally for principal components).

| Study | Number of Cases | Number of Controls | Per 5 unfavorable average alleles | | | High versus low GRS group | | | | | |
| | | | | | | 5% versus 95% | | | 10% versus 90% | | |
			OR	95% CI	P (1-sided)	OR	95% CI	P (1-sided)	OR	95% CI	P (1-sided)
(A) ESKD (cases: ICD10 code N18.0 or N18.5; controls: no ICD10 code N18, eGFR>60 mL/min/1.73m², frequency-matched by age-group and sex)											
4D_KORA-F3	1,100	1,601	1.122	0.925,1.362	0.121	1.260	0.669,2.377	0.237	1.526	0.978,2.379	0.0313
GENDIAN_KORA-F4	470	1,545	1.146	0.923,1.423	0.108	0.954	0.468,1.946	0.449	1.036	0.625,1.719	0.445
UKBBCa_co	498	1,494	1.085	0.885,1.330	0.216	1.220	0.639,2.329	0.273	1.479	0.921,2.373	0.0525
Meta-analysis	**2,068**	**4,640**	1.117	0.993,1.256	**0.0329**	1.150	0.785,1.686	0.236	1.349	1.027,1.773	0.0157
(B) AKI (cases: ICD 10 code N17; controls: no ICD10 code N17, eGFR>60 mL/min/1.73m², frequency-matched by age-group and sex)											
UKBBCaCo	3,878	11,634	1.179	1.095,1.270	6.47×10^{-06}	1.524	1.204,1.931	4.70×10^{-04}	1.272	1.080,1.499	1.97×10^{-03}

Study=Study name, **OR**=Odds Ratio of the GRS-association, **95% CI**=95% confidence interval of the association, **P (1-sided)**=1-sided association P-value, **ESKD**=End-stage Kidney Disease, Individuals analyzed here are distinct from the eGFR-decline GWAS except for the KORA-F3 and KORA-F4 controls. **AKI**=Acute Kidney Injury, UKBBCaCo=cases and controls from UK Biobank distinct from UK Biobank study participants used in the GWAS for eGFR decline.

Figure 1: Eleven loci identified by GWAS for eGFR-decline unadjusted and/or adjusted for eGFR-baseline. We conducted GWAS for eGFR-decline baseline-unadjusted and baseline-adjusted (n up to 343,339 or 320,737, respectively). Shown are association P-values versus genomic position, identified loci annotated by nearest gene: (A) association for eGFR-decline baseline-<u>unadjusted</u> identified one genome-wide significant locus for decline (P<5x10^{-8}) and two Bonferroni-corrected significant loci among the 263 lead variants for cross-sectional eGFR[15] outside of *UMOD-PDILT* (red dots, P<0.05/263=1.90x10^{-4}; known locus for decline marked in blue; novel loci for this phenotype in orange); (B) association for eGFR-decline baseline-<u>adjusted</u> identified 8 additional loci (novel loci marked in green; known loci or loci already identified in (A) marked in blue). Altogether, 11 loci were identified with genome-wide significance for eGFR-decline unadjusted and/or adjusted for eGFR-baseline.

Figure 2: Relationship of SNP-effects on eGFR-decline baseline-unadjusted with baseline-adjusted effects for the 12 identified variants. Shown are: **(A)** SNP-effects per year and allele for eGFR-decline baseline-<u>unadjusted</u> ("decline") versus eGFR-decline baseline-<u>adjusted</u> in all studies ($n_{decline}$=343,339; $n_{decline-adj}$=320,737) and restricted to studies where baseline-adjusted results were computed rather than formula-derived (inserted panel, n=103,970); red line indicates identify line); **(B)** standardized SNP-effects per year and allele for eGFR-decline baseline-<u>unadjusted</u> ($\hat{\beta}_{DECLINE}/sd_{DECLINE}$, n=343,339) and per allele for cross-sectional eGFR on ln-scale ($\hat{\beta}_{BL}/sd_{BL}$, n=765,348 [15]); grey line indicates phenotype correlation line y=0.34*x (0.34=mean phenotype correlation across studies). For A&B: coding allele is the faster-decline allele (=cross-sectional eGFR-lowering allele). Color codes whether SNP was identified for decline baseline-unadjusted and/or baseline-adjusted. **(C)** Illustration of the SNP-effect for eGFR-decline baseline-adjusted (standardized to Y-scale) as a sum of the SNP-effect baseline-unadjusted (standardized) and the correlation-weighted SNP-effect on eGFR at baseline (standardized).

Figure 3: Relationship of SNP-by-age interaction effects for eGFRcys with those of eGFRcrea and with SNP-effects for eGFR-decline for the 12 identified variants. Shown are SNP-by-age interaction effect sizes per year and allele for cross-sectional eGFRcys (UK Biobank individuals independent from GWAS, $n_{SNPxage}$=351,601; main age effect modelled non-linearly, main SNP-effect linearly, age effect and SNP effect in interaction term linearly, age centered at 50 years) versus: **(A)** SNP-by-age interaction effects on cross-sectional eGFRcrea ($n_{SNPxage}$=351,462), **(B)** SNP-effects on eGFR-decline baseline-unadjusted per year and allele ($n_{decline}$=343,339). Coding allele is the faster-decline allele (=cross-sectional eGFR-lowering allele); color code as in **Figure 2**; red line indicates identity line; symbol types code significance of interaction term (P< 0.05/12). Among the 9 SNPs with genuine eGFR-decline association, 7 SNPs showed interaction for eGFRcrea or eGFRcys (all negative), and all 3 SNPs without genuine eGFR-decline association showed no interaction for eGFRcys (one with positive significant interaction for eGFRcrea).
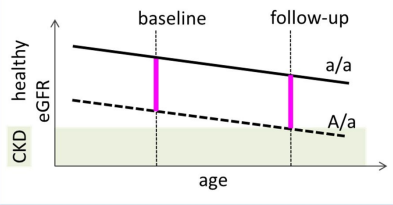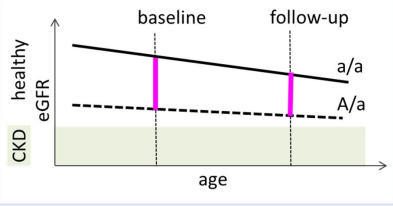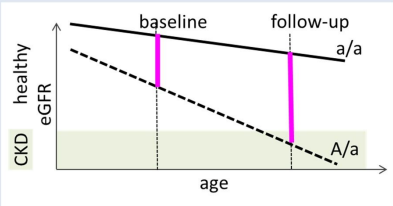
Figure 4: A concept for three classes of SNP-associations on cross-sectional eGFR distinguished by the presence and direction of the SNP-association with eGFR-decline. Let *A/a* be the genotype group of individuals with, on average, lower cross-sectional eGFR compared to *a/a* (*A*=effect allele). Let's further assume that eGFR-declines monotonously by age (approximated as linear decline) and that there is no "cross-over" between genotype groups. Shown are (left) a graphical scheme, (middle) the theoretical association, (right) the observed SNPs in line with the respective class. In the three graphical schemes, **black** lines illustrate mean eGFR-decline by genotype group; SNP-effects on eGFR for these individuals captured cross-sectionally at different ages are magenta. When a cross-sectional study captures individuals of relevant ages, the SNP-effects on eGFR should show an interaction by age for *class II* and *class III* SNPs (positive and negative, respectively). The 9 variants with genuine eGFR-decline association were *class III*, while the other 3 variants were *class I*.

Figure 5: Data, analyses, and results in a nutshell.

26

A – eGFRcrea decline

B – eGFRcrea decline_adj

	theory	observed
Class I	Association with baseline eGFR (intercept), but not with eGFR-decline (slope); i.e. stable SNP-effect on eGFR when individuals get older.	3 SNPs: near *SHROOM3, CPS1, GATM;* no SNP-by-age interaction for eGFRcys
Class II	Association with baseline eGFR (intercept) AND eGFR-decline with a flatter slope for A/a; i. e. diminishing SNP-effect on eGFR by age.	No SNP
Class III	Association with baseline eGFR (intercept) AND eGFR-decline with a steeper slope for A/a; i. e. increasing SNP-effect on eGFR by age.	9 SNPs: near *UMOD* (2), *PRKAG2, SPATA7, FGF5, OVOL1, TPPP, C15ORF54, ACVR2B;* 7 with SNP-by-age interaction for eGFRcys

A = effect allele as the allele associated with lower baseline eGFR   decline = unadjusted for eGFR-baseline

Data

Annual eGFR-decline unadjusted for eGFR-baseline	Annual eGFR-decline adjusted for eGFR-baseline eGFR
CKDGen and UK Biobank N = 343,339 (62 studies)	CKDGen and UK Biobank N = 320,737 (50 studies)

GWAS

12 variants (11 loci) associated genome-wide significantly ($P<5\times10^{-8}$) for eGFR-decline unadjusted and/or adjusted for eGFR-baseline (10 loci novel for eGFR-decline, 1 known; all 11 loci known for cross-sectional eGFR)

9 variants genuinely associated for eGFR-decline (i.e. Bonferroni-corrected significance for eGFR-decline unadjusted for eGFR-baseline, $P<0.05/12=4.17\times10^{-3}$)

3 variants with zero effect on eGFR-decline unadjusted for eGFR-baseline

UMOD-PDILT (2x)* SPATA7* PRKAG2* FGF5* OVOL1* TPPP* C15ORF54 ACVR2B

GATM CPS1 SHROOM3

*also significant SNP x age interaction for cross-sectional eGFRcys

Two observed of three theoretical classes of variants associated with cross-sectional eGFR

Class III: association of with eGFR-decline (slope) with a steeper slope for the cross-sectional eGFR-lowering allele

Class II: association with slope (flatter slope for eGFR-lowering allele)

Class I: no association with slope

Genetic effect sizes in high-risk groups

2- to 4-fold larger eGFR-decline effects in individuals with DM or CKD at baseline (5 of 9 variants Bonferroni-corrected significant for CKD progression, $P<0.05/9=5.56\times10^{-3}$)

Candidate genes for eGFR-decline and CKD progression

5 genes (4 novel, UMOD known) mapped by identified eGFR-decline variants via modulated gene expression in kidney tissue (eQTL)

UMOD § SPATA7 § GALNTL5 (in PRKAG2 locus) § TPPP § FGF5   §also for CKD progression

9-variant GRS association with ESKD and AKI

OR of 1.35 and 1.27 (95% CI= 1.03-1.77 and 1.08-1.50) for ESKD and AKI, respectively, comparing individuals with high versus low genetic risk (90th versus 10th percentile of GRS)

Abbreviations: DM=diabetes mellitus, CKD = chronic kidney disease, CKD progression = eGFR-decline in CKD individuals, eQTL = expression quantitative trait locus variant, GRS = genetic risk score, CI= confidence interval, ESKD= end-stage kidney disease, AKI = acute kidney injury.

SUPPLEMENTARY ONLINE MATERIAL

Genetic loci and prioritized genes for kidney function decline

from a meta-analysis of 62 longitudinal genome-wide association studies

Supplementary Methods

Supplementary Notes

Supplementary Figures

Supplementary Tables

Extended acknowledgements, study funding information and author contributions

Supplementary Methods

General approach for GWAS meta-analysis

An analysis plan and standardized scripts for phenotype generation and GWAS analyses were developed and implemented in all 61 CKDGen studies and UK Biobank. The 61 CKDGen studies consisted of 58 studies that were long-term partners of CKDGen ("old" CKDGen studies) and three studies that have joined CKDGen more recently allowing for more elaborate analyses (AugUR, HUNT and MGI; extended analysis plan, see below). Most studies were population-based and thus including individuals with specific kidney diseases according to the prevalence in the general population. Each study conducted GWAS analyses according to this pre-defined plan, separately by ancestry (if applicable). Ancestry was defined by genetic principal components or participants' self-report. For each study, phenotypic information and genome-wide summary statistics per SNP were transferred to the meta-analysis centers.

Each study had been conducted according to the declaration of Helsinki. The studies have been approved by each local ethics committee. All participants in all studies provided written informed consent.

Meta-analyses were conducted, significant variants identified and respective locus regions selected. A GWAS across all available studies was shown to be advantageous over conducting a discovery followed by a replication stage on selected variants[S1,S2]. Therefore, rather than conducting a discovery GWAS in old CKDGen studies and a replication in recently joined CKDGen studies and UK Biobank, we included all studies into the GWAS meta-analysis on eGFR decline.

Phenotype definition

In each contributing study, serum creatinine was measured at least two times, utilizing two measurements at largest time distance (study-specific details in **Supplementary Table S1**). When measurements were obtained by Jaffé assay (before 2009), creatinine measurements were calibrated (multiplying by 0.95[S3]). Serum creatinine measured at baseline and follow-up was used to estimate eGFR at baseline and follow-up, respectively, according to the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation[S4]. This equation contains an age, sex, and ancestry term for a best fit of creatinine-based eGFR to measured GFR. At baseline and follow-up, eGFR was winsorized at 15 and 200 mL/min/1.73m². Annual eGFR-decline was defined as "- (eGFR at follow-up - eGFR at baseline) / number of years of follow-up"; thus, eGFR-decline is positive when eGFR is lower at follow-up compared to baseline and comparable across studies with different follow-up length.

In each study, eGFR-decline was analyzed overall and restricted to individuals with CKD or DM at baseline. CKD at baseline was defined as eGFR<60 mL/min/1.73m² at baseline. In CKDGen, DM at baseline was defined as fasting plasma glucose ≥126 mg/dl (7.0 mmol/L)

or diabetes therapy, or (fasting glucose unavailable) as non-fasting plasma glucose ≥200 mg/dl (11.0 mmol/L) or diabetes therapy, or (glucose unavailable) as self-reported diabetes. For UK Biobank, DM was defined as HbA1c≥48 mmol/mol or diabetes therapy.

Study-specific generation of outcome variables according different adjustment models

In each study, different models for the SNP-association with annual eGFR-decline as outcome were computed genome-wide: (i) adjusted for age, sex, and DM and applied to all individuals ("decline DM-adjusted"); (ii) adjusted for age and sex restricted to individuals with DM or CKD at baseline ("decline in DM", "decline in CKD"). In the recently joined CKDGen studies and UK Biobank, an extended suite of models was applied: additional analyses were (iii) adjusted for age and sex using all individuals ("decline"), (vi) adjusted for age, sex and eGFR baseline using all individuals (eGFR baseline on log-scale, ln(eGFR), "decline adjusted for baseline"). Further study-specific adjustments were applied (as applicable), including genetic principal components to account for population substructure.

These adjustments were implemented by generating residuals of annual eGFR-decline adjusted for the respective covariates and using these residuals as outcome in GWAS. This is a standard approach yielding comparable results to using the unadjusted phenotype as outcome in GWAS adjusting for the respective covariates. The utilized approach implies fewer covariates in GWAS being computationally more efficient. We standardized the creation of these outcome variables for GWAS by providing a centrally developed script, which also provided descriptive statistics on the study-specific phenotype.

Genotyping, imputation, and study-specific GWAS

In each study, genotyping was conducted using Affymetrix and Illumina arrays (**Supplementary Table S2**). Imputation was performed using 1000 Genomes[S5] phase 1 or phase 3, the Haplotype Reference Consortium[S6] v1.1 or customized reference panels, annotating all variants on the GRCh build 37 reference build; imputed genotypes were coded as allelic dosages and imputation quality was provided as IMPUTE2[S7] info score, MACH/minimac[S8] RSQ or similar; quality control before and after imputation was conducted study-specifically (**Supplementary Table S2**).

In each study, GWAS analyses were conducted according to the centrally defined analysis plan. CKDGen studies included different ancestries (European, African American, East Asian, South Asian, and Hispanic) and contributed analyses ancestry-specific. Since most CKDGen studies individuals were European ancestry (94.90%), UK Biobank analyses focused on unrelated European ancestry individuals where two assessments of eGFR were available (n=15,442). For each GWAS, linear regression on the respective outcome variable was computed per SNP (modelled as allele dosages linearly) adjusted for principle

components and other study-specific covariates as applicable (**Supplementary Table S2**). This yielded three GWAS results for "old" and recently joined CKDGen studies (decline DM-adjusted, decline in DM, decline in CKD) and two further GWAS results for recently joined CKDGen studies and UK Biobank (decline, decline adjusted for baseline). Summary statistics were collected and quality controlled centrally with GWAtoolbox[S9].

Study-specific summary statistics for decline adjusted for baseline

As noted above, GWAS results on eGFR-decline adjusted for eGFR-baseline was not available in all studies. GWAS meta-analyses logistics in so many studies are highly complex; it is not trivial to "add" analyses applying other models. However, there is mathematical help to facilitate covariate adjustment post-hoc, i.e. by formula, based on GWAS summary statistics unadjusted for eGFR-baseline and GWAS summary statistics for eGFR-baseline and study-specific phenotype information[S10]. We demonstrate how this works (**Supplementary Note S1**) and that it works in this setting by validation studies: we compared formula-derived summary statistics for baseline-adjusted decline with model-computed baseline-adjusted decline in a subset of studies (the recently joined CKDGen studies, UK Biobank, selected "old" CKDGen studies). For eGFR-decline adjusted for baseline in the following, we used formula-derived summary statistics for the "old" CKDGen studies and computed summary statistics for the recently joined studies and UK Biobank.

Meta-analyses of GWAS summary statistics

Before meta-analysis, we excluded, from each study file, multi-allelic variants, variants with a Minor Allele Count <10, and variants with an imputation quality <0.6 (R^2 from minimac[S8] or info score from Impute[S7]). Per study, genomic control (GC) correction was applied when the GC-factor lambda was >1. We excluded a study for a specific analysis, when it contributed <100 individuals after quality control for this analysis.

Per model, we conducted a fixed-effects inverse-variance-weighted meta-analysis using metal[S11]. To account for the sequential recruitment of studies, we meta-analyzed per-variant summary statistics across "old" CKDGen studies (GC-corrected) and across recently joined CKDGen studies plus UK Biobank (GC-corrected), and then meta-analyzed these two (again GC-corrected, **Supplementary Figure S1**). After meta-analysis, only variants present in ≥50% of GWAS files and minor allele frequency ≥1% were retained for further analyses.

Identification of associated loci

For our GWAS search, we selected genome-wide significant variants ($P<5.00\times10^{-8}$) in the meta-analyzed summary statistics and identified independent locus lead variants by an iterative approach, as applied previously[S12]: (i) from all genome-wide significant variants, we

selected the variant with the smallest P-value as the first lead variant and defined this variant's locus region as lead variant ±500kB, (ii) omitting this identified region, we selected the next variant with the smallest P-value, and (iii) repeated this procedure until no further variant with P-value<5.00×10^{-8} was observed. The *MHC* region (chr6:28.5-33.5MB) was considered a single locus. We checked for overlapping loci, but there were none.

For the candidate-based approach, we used the 265 lead variants previously reported for association with cross-sectional ln(eGFR)[S12], excluded the locus regions identified by the GWAS search, and, for the remaining candidate variants, judged significance at Bonferroni-corrected level.

For identified variants, we evaluated ancestry-related heterogeneity using MR-MEGA v.0.1.5 (Meta-Regression of MultiEthnic Genetic Association[S13], including three principle components. We also conducted sensitivity analyses incorporating further models of covariate adjustment for identified eGFR-decline associations in a validation meta-analysis.

SNP-by-age interaction on cross-sectional eGFR

We investigated the lead SNPs identified for (creatinine-based) eGFR-decline for SNP-by-age interaction on cross-sectional eGFR (based on creatinine or cystatin C, eGFRcrea, eGFRcys). For this we used data that was independent of the SNP identification step: unrelated European ancestry UK Biobank individuals with one eGFRcrea or eGFRcys assessment excluding the 15,442 individuals in the decline GWAS (yielding > 350,000 individuals).

For each SNP, we applied two linear regression models, one each for the outcome eGFRcrea or eGFRcys, using the covariates age, sex, SNP, SNP-by-age interaction term, and four principal components (age centered at 50 years). We modelled (i) the main age effect on the outcome allowing for non-linear effects (to avoid spurious effects from non-linear main age effect when modelling age linearly), (ii) the main SNP effect linearly per allele dosage, and (iii) for the SNP-by-age interaction effect, the SNP-effect was modelled linearly per allele dosage and the age effect was allowed to vary non-linearly (smooth function, varying coefficient model[S14], penalized thin-plate regression splines, mgcv-package in R[S15]). In a second analysis, the age effect in the SNP-by-age interaction was modelled linearly (i.e. linear effects for both SNP and age in the SNP-by-age term). We judged significance of the interaction at Bonferroni-corrected level.

Genetic effect sizes and GRS analysis for eGFR-decline

We provide SNP-specific effect sizes on eGFR-decline in mL/min/$1.73m^2$ per year over all individuals and focused on individuals with DM at baseline or CKD at baseline. We provide cumulative effects by GRS analysis in the population-based study HUNT (19-90 years old, European ancestry, up to 21 years of follow-up, mean of age-/sex-adjusted residuals for eGFR-

decline = 1.02 mL/min/1.73m^2/year). To compute the GRS, we counted the number of the faster-decline allele across identified variants for each study participant, weighted by the effect size for eGFR-decline unadjusted for eGFR-baseline, then divided by the sum of weights and multiplied by the number of variants in the GRS. By this, the GRS is scaled from 0 to 2 times the number of variants, where one unit reflects one "average" unfavorable allele. We tested the quantitative GRS with eGFR-decline via linear regression adjusted for age and sex (unadjusted for eGFR-baseline) and we compared individuals with high versus low GRS (≥95th versus ≤5th percentile, ≥90th versus ≤10th percentile derived from UK Biobank excluding individuals in the eGFR-decline GWAS). This was done over all individuals and restricted to individuals with DM at baseline or CKD at baseline.

We also computed a SNP's genetic effect size relative to the phenotype variance as beta-estimates^2 * Var(SNP) / Var(Y), i.e. beta-estimates^2 * 2*MAF*(1-MAF)/ (standard deviation of Y)^2, where MAF is the minor allele frequency of the respective variant. The joint effect of several variants was derived as the sum of the respective SNPs' effects. For this, again, we used the phenotype variance from HUNT: the standard deviation of age-/sex-adjusted residuals for eGFR-decline = 0.91 mL/min/1.73m^2 overall, 1.25 mL/min/1.73m^2 among individuals with DM, 1.39 mL/min/1.73m^2 with CKD, and for eGFR cross-sectional =0.12 mL/min/1.73m^2 on the log-scale.

GRS analyses for ESKD and AKI

We were interested in whether the GRS across the variants identified for eGFR-decline showed association with severe clinical endpoints, ESKD and AKI. For this, we used three case sets for ESKD and one case set for AKI as well as controls (eGFR>60 mL/min/1.73m^2) from population-based studies frequency-matched with regard to age-group and sex as described previously[S16]. Briefly, the three ESKD studies consisted of: (i) ESKD cases from unrelated European ancestry UK Biobank individuals (ICD10 code N18.0 or N18.5, i.e. need for dialysis) and matched UK Biobank controls (no record of any N18), excluding individuals in eGFR-decline GWAS; (ii) ESKD cases from GENDIAN and controls from KORA-F4; (iii) ESKD cases from the 4D-study[S17] and controls from KORA-F3. The study on AKI used AKI cases from UK Biobank (ICD10 code N17, "Acute Renal Failure") and UK Biobank controls (no record of N17), excluding individuals in eGFR-decline GWAS. By this, the cases and controls across all four studies were independent of eGFR-decline GWAS, except the KORA-F3 and KORA-F4 controls to keep the previously designed and published case-control comparisons with GENDIAN and 4D.

For each of these four case-control studies, we retrieved the respective SNPs and computed a weighted GRS across identified variants for each individual as described above. We tested the quantitative GRS with ESKD or AKI. We applied a one-sided test, since we were

7

only interested in this association when the GRS increased the odds of ESKD or AKI. We also compared individuals with high versus low GRS (≥95th GRS percentile, ≤5th percentile and ≥90th versus ≤10th GRS percentile, defined in UK Biobank individuals excluding individuals in eGFR-decline GWAS) and tested (one-sided) for increased odds of ESKD (meta-analysis across the three studies) or AKI. Associations are derived via logistic regression adjusted for matching variables age-groups and sex (for AKI additionally for the first two principal components).

Supplementary Note S1: Equivalence of DM-adjusted versus not DM-adjusted GWAS on eGFR-decline in the validation meta-analysis

In the recently joined studies (HUNT, MGI, AugUR) and UK Biobank, we had more adjustment models computed for GWAS on eGFR-decline, to better understand similarities and differences. In these, we compared the GWAS summary statistics for eGFR-decline adjusted for DM-status to GWAS without adjustment for DM-status (i.e. GWAS on age- and sex-adjusted residuals and with and without adjustment for DM-status at baseline). In each study, we found precisely the same beta-estimates and standard errors (SE): (i) for the 265 SNPs identified previously for cross-sectional eGFR[S12], for which we had a prior hypothesis that these contained the SNPs associated with eGFR-decline, as well as (ii) genome-wide where most of the SNP-associations are under the Null (**Supplementary Figure S4A**).

We added further "old" CKDGen studies to substantiate these findings in further studies and in an expanded validation meta-analysis (n=103,970). Again, we found DM-adjusted and not DM-adjusted beta-estimates and SEs to be precisely the same (**Supplementary Figure S4A**). Of note, this validation meta-analysis included general population studies and studies of specific scope: hospital-based (MGI), focused on individuals aged 70+ years (AugUR), or focused on individuals with chronic kidney disease (GCKD).

Given this equivalence, we did not distinguish any more between results DM-unadjusted or DM-adjusted.

Supplementary Note S2: Formula-based covariate adjustment using GWAS summary statistics

Let's assume we have a quantitative phenotype Y and a covariate C. Let's further assume, we have GWAS summary statistics as beta-estimates and respective standard errors, $\hat{\beta}_Y$ and $\widehat{SE}_Y$ (beta-estimate and standard error) from linear regression models per genetic variant, i.e. from $Y \sim \alpha + \beta_Y SNP$ (unadjusted model, omitting the indexing per variant). Let's assume we also have GWAS summary statistic $\hat{\beta}_C$ and $\widehat{SE}_C$ for the covariate C from the model $C \sim \alpha + \beta_C SNP$ (covariate model via linear regression, C binary or quantitative). We can then "adjust" the summary statistics formula-based, i.e. we can derive the GWAS summary statistics $\hat{\beta}_{YadjC}$ and $\widehat{SE}_{YadjC}$ for the adjusted model, $Y_{adjc} \sim \alpha + \beta_{YadjC} SNP + yC$, as described[S18] according to

$$\hat{\beta}_{YadjC} = \hat{\beta}_Y - \left(r_{YC} * \frac{sd_Y}{sd_C} \right) * \hat{\beta}_C \text{ and}$$

$$\widehat{SE}_{YadjC} = \sqrt{\widehat{SE}_Y^2 + \left(r_{YC} * \frac{sd_Y}{sd_C} \right)^2 * \widehat{SE}_C^2 - 2 * \left(r_{YZ} * \frac{sd_Y}{sd_C} \right) * corr(\hat{\beta}_Y, \hat{\beta}_C) * \widehat{SE}_Y * \widehat{SE}_C}.$$

Here, we assume that we know the standard deviation of C and Y, sd_C and sd_Y, respectively, the phenotypic correlation, r_{YC} (estimated as Pearson correlation coefficient between Y and C) and the genetic correlation between Y and C, $corr(\hat{\beta}_Y, \hat{\beta}_C)$, (using all genetic effects for Y and C genome-wide for estimation as reasonable proxy). When r_{YC} is zero, the adjusted model SNP-effects, $\hat{\beta}_{YadjC}$, are the same as the unadjusted model SNP-effects, $\hat{\beta}_Y$.

Alternatively, when we have GWAS summary statistics from the adjusted model, $Y_{adjc} \sim \alpha + \beta_{YadjC} SNP + yC$, and the covariate model, $C \sim \alpha + \beta_C SNP$, we can "de-adjust" summary statistics formula-based, i.e. we can derive the GWAS summary statistics of the unadjusted model as

$$\hat{\beta}_Y = \hat{\beta}_{YadjC} + \left(r_{YC} * \frac{sd_Y}{sd_C} \right) * \hat{\beta}_C \text{ and}$$

$$\widehat{SE}_Y = \sqrt{\widehat{SE}_{YadjC}^2 + \left(r_{YC} * \frac{sd_Y}{sd_C} \right)^2 * \widehat{SE}_C^2 + 2 * \left(r_{YZ} * \frac{sd_Y}{sd_C} \right) * corr(\hat{\beta}_{YadjC}, \hat{\beta}_C) * \widehat{SE}_{YadjC} * \widehat{SE}_C}$$

.

We apply this on our example to summary statistics for annual eGFR-decline adjusted for eGFR-baseline (BL): given the beta-estimates for decline unadjusted for ln(eGFRcrea_BL) (in fact, residuals adjusted for age, sex), $\hat{\beta}_{decline}$, and the beta-estimates for ln(eGFRcrea_BL) (i.e. residuals adjusted for age and sex), $\hat{\beta}_{BL}$, we can "adjust" results for BL using the formula, i.e., derive the beta-estimates for decline adjusted for BL (residuals adjusted for age and sex), $\hat{\beta}_{decline\_adj\_BL}$, as

$$\hat{\beta}_{decline\_adj\_BL} = \hat{\beta}_{decline} - \left(r_{decline,BL} * \frac{sd_{decline}}{sd_{BL}} \right) * \hat{\beta}_{BL}.$$

Effect sizes here are given for the BL-lowering effect allele (which is usually the decline-increasing allele). The can also be written as

$$\frac{\hat{\beta}_{decline\_adj\_BL}}{sd_{decline}} = \frac{\hat{\beta}_{decline}}{sd_{decline}} + r_{decline,BL} * \left(-\frac{\hat{\beta}_{BL}}{sd_{BL}}\right).$$

This shows that the effect size of decline adjusted for BL standardized to the scale of standardized $\hat{\beta}_Y$ effects (i.e. divided by $sd_{decline}$) is the sum of (i) the (standardized) effect size of decline unadjusted (i.e. the vertical distance of this effect to the x-axis in a $\hat{\beta}_Y/sd_Y$ versus $\hat{\beta}_C/sd_C$ plane) and (ii) the vertical distance from the intersection point of the x-axis at $\hat{\beta}_C/sd_C$ (i.e. < 0 when the coding allele is the $\hat{\beta}_C$-lowering allele) to the phenotype correlation line, $f(x) = r_{YC} * x$, when the phenotype correlation is positive, like r_{YC}=0.33 in UK Biobank, i.e. to the point ($\hat{\beta}_C/sd_C$, 0.33*$\hat{\beta}_C/sd_C$). This also shows that $\hat{\beta}_{decline_{adj_{BL}}} < \hat{\beta}_{decline}$, since $\hat{\beta}_C < 0$, by definition.

Supplementary Note S3: Validation of the formula-derived association for eGFR-decline adjusted for eGFR-baseline

In the recently joined studies and UK Biobank, we had more adjustment models computed for GWAS on eGFR-decline, to better understand similarities and differences. In these, we compared the summary statistics for eGFR-decline adjusted for eGFR-baseline (i.e. age- and sex-adjusted residuals and additional adjusted for ln(eGFRcrea baseline)) with eGFR-decline unadjusted for eGFR-baseline (i.e. age- and sex-adjusted residuals) and found substantial differences (**Supplementary Figure S4B**). Thus, the two models, unadjusted and adjusted for eGFR-decline were considered further.

Generally, in GWAS meta-analysis, the number of GWAS models computed needs to be as parsimonious as possible to remain feasible. In each of the "old" CKDGen studies, we had GWAS summary statistics for eGFR-decline unadjusted for eGFR-baseline, GWAS summary statistics for cross-sectional eGFR, and study-specific phenotypic information. We knew that this enabled us to do the adjustment by formula[S10,S18] (**Supplementary Note S1**). For the "old" CKDGen studies, we thus derived GWAS summary statistics for eGFR-decline adjusted for eGFR-baseline applying this formula.

While the formula was established previously[S10], we validated that it worked in this setting using the recently joined CKDGen studies and UK Biobank, where we had the model "eGFR-decline adjusted for eGFR-baseline" computed: we also derived the SNP-associations for "eGFR-decline adjusted for eGFR-baseline" based on the formula for comparison in these studies for the purpose of validation. We found the formula to work very precisely per study: we observed equivalence in beta estimates and SEs when focused on the 265 SNPs identified previously for cross-sectional eGFR[S12], for which we had a prior hypothesis that these contained the SNPs associated with eGFR-decline, as well as genome-wide, where most SNP-

associations were under the Null (**Supplementary Figure S4C**; e.g., in UK Biobank for the 265 variants: Pearson correlation coefficient r=1.00 for betas and SEs; maximum difference in beta=3.26×10^{-2}, maximum differences in SEs =1.01×10^{-3}). We added further "old" CKDGen studies also to yield an expanded validation meta-analysis (n=103,970). Again, we found the formula to work precisely in each study and in the expanded validation meta-analysis (**Supplementary Figure S4C**).

The formula is mathematically derived and works perfectly when GWAS summary statistics for baseline eGFR are available. For studies with GWAS on cross-sectional eGFR, the sample size for cross-sectional eGFR is typical a bit larger than the sample size for eGFR-baseline for longitudinal studies (i.e. restricting to individuals in the follow-up). We evaluated the impact of using cross-sectional eGFR summary statistics rather than baseline eGFR summary statistics in the formula in three "old" CKDGen studies at the hand of the Regensburg meta-analysis center. There was no difference in SEs for the 265 variants or genome-wide, a slight difference for beta estimates of the 265 variants, and a larger (random, not biased) difference in betas genome-wide (**Supplementary Figure S4D**). This difference in genome-wide SNP-estimates can be attributed to random noise in the per-variant estimates under the null hypothesis (considering most genome-wide SNPs as not associated with eGFR-decline). We extended this validation experiment by three further studies, and found the same (**Supplementary Figure S4D**). In summary, we concluded that the formula-derived association estimates worked well in this setting for the 265 variants and also, with some more random noise, for the other genome-wide variants.

Of note, these validation meta-analyses included general population studies as well as studies of specific scope: hospital-based (MGI), focused on individuals aged 70+ years (AugUR), focused on individuals with chronic kidney disease (GCKD), or focused on individuals with DM (Diacore).

Supplementary Note S4: Graphical illustration of the relationship between SNP-effects on eGFR-decline unadjusted and adjusted for eGFR-baseline.

Figure 2C provides an informative geometrical illustration for the relationship between a SNP-effect on eGFR-decline baseline-unadjusted (standardized, depicted on Y-axis), $\hat{\beta}_{DECLINE}/sd_{DECLINE}$, and the SNP-effect on eGFR-decline baseline-adjusted (standardized to Y-axis scale), $\hat{\beta}_{DECLINE\_adj\_BL}/sd_{DECLINE} = \hat{\beta}_{DECLINE}/sd_{DECLINE} + r_{decline,BL} * (-\hat{\beta}_{BL}/sd_{BL})$, where $r_{DECLINE,BL}$ is the phenotypic correlation of baseline-unadjusted eGFR-decline with baseline eGFR and $\hat{\beta}_{BL}/sd_{BL}$ is the standardized variant effect on baseline eGFR.

While this relationship was derived per study (**Supplementary Note S1**), this also holds approximately for meta-analyzed effect sizes, as mostly the same studies contributed to the respective meta-analyses. The difference between the two effects, baseline-adjusted and

baseline-unadjusted decline, $r_{decline,BL} * (-\hat{\beta}_{BL}/sd_{BL})$, can be visualized when adding the phenotype correlation line, $f(x) = r_{DECLINE,BL} * x$ (mean correlation across studies= 0.34): while the baseline-unadjusted decline effect, $\hat{\beta}_{DECLINE}/sd_{DECLINE}$, is the vertical distance from symbol to X-axis, the baseline-adjusted decline effect, $\hat{\beta}_{DECLINE\_adj\_BL}/sd_{DECLINE}$, is the vertical distance from symbol to phenotype correlation line.

Supplementary Note S5: Comparison of the signals for eGFR-decline unadjusted and adjusted for eGFR-baseline and cross-sectional eGFR for the 11 identified loci

We compared the association signals for the 11 identified loci for eGFR-decline (unadjusted for eGFR-baseline) with signals for eGFR-decline adjusted for eGFR-baseline with signals for eGFR cross-sectional[S12] in regional association plots (**Supplementary Figure S5A-C**),

For the 4 variants identified for eGFR-decline unadjusted for eGFR-baseline, we found unadjusted eGFR-decline signals to coincide with adjusted eGFR-decline signals and with cross-sectional eGFR signals (**Supplementary Figure S5A**). Lead variants for unadjusted eGFR-decline (i.e. the variant with the smallest P-value for unadjusted eGFR-decline) were the same or highly correlated with the respective cross-sectional lead variants (r²=same, same, 1.00 and 0.93 for *UMOD-PDILT* (2), *PRKAG2* and *SPATA7*, respectively).

Among the 5 lead variants identified by GWAS on eGFR-decline adjusted for eGFR-baseline with significant association for eGFR-decline unadjusted for eGFR-baseline (i.e. "genuine" eGFR-decline variants, **Supplementary Figure S5B**), all signals for decline adjusted coincided with respective signals for decline unadjusted, except for the *TPPP* locus (but there, the signal for decline unadjusted sharpened when including the studies with lower imputation quality and then coincided). Three of the 5 lead variants were the same as (*FGF5*) or highly correlated with (*C15ORF54* and *ACVR2B,* R²=0.61 and 0.98) the respective lead variants for decline unadjusted. In the *OVOL1* locus, the lead variant for decline adjusted (rs4930319) depicted the same association signal as for decline unadjusted, but was not highly correlated with the variant with the smallest P-value for decline unadjusted (R² with rs117829045=0.11) due to differing allele frequencies (MAF=0.11 and 0.33, respectively); the variants were suggested to be inherited via the same haplotypes (D'=1.00). Among the 5 variants, we found 3 signals for eGFR-decline adjusted for eGFR-baseline to coincide with the signal for cross-sectional eGFR (for *FGF5*, *OVOL1*, *ACVR2B)* and lead variants for decline adjusted as highly correlated with the respective lead variants for cross-sectional eGFR (r²= 0.95, 0.98, 0.96, respectively; **Supplementary Figure S5B**). In *C15ORF54* and *TPPP* loci, the decline adjusted signal appeared to be a 2nd signal for cross-sectional eGFR: the lead variant for decline adjusted were not correlated with the lead variant for cross-sectional eGFR (R²= 0.04 and 0.11). The lead variant for decline adjusted near *TPPP* depicted a cross-sectional signal 22kb distant from the reported cross-sectional lead variant with different allele

frequencies (MAF=0.49 and 0.27, respectively; D'=0.57); of note, the lead variants for decline adjusted captured a 2nd signal identified in the recently published cross-sectional eGFR analysis[S19] and there the lead variants were exactly the same. The *C15ORF54* lead variant for decline adjusted was highly correlated with a 2nd signal for cross-sectional eGFR (rs28833881, r^2=0.98).

For the 3 loci identified by eGFR-decline adjusted for eGFR-baseline without significant association with eGFR-decline unadjusted for eGFR-baseline (i.e., not a genuine eGFR-decline association), there was no signal for decline unadjusted (*GATM*, *CPS1*, *SHROOM3*; **Supplementary Figure S5C**). The lead variants for decline adjusted were the same or highly correlated with the respective cross-sectional eGFR lead variant (R^2=0.98, same, 0.59).

Supplementary Note S6: Age-dependency of SNP-effects and main age effect on eGFR.

Before interpreting SNP-by-age interaction effects on cross-sectional eGFRcrea and eGFRcys, we evaluated the main age effect on eGFRcrea and eGFRcys (i.e. age and sex in the model). We found large main age effects, which were fairly linear: beta-estimate per year of age [95%-CI] = -0.775 units [-0.780, -0.771] and -1.024, [-1.030, -1.019] on eGFRcrea or eGFRcys, respectively (**Supplementary Figure S6Z**). We nevertheless allowed for non-linear main age effects in the SNP-by-age interaction analyses, since the main age effect was large and even a slight deviation from non-linearity can distort interaction effects if unaccounted.

We found the age-dependency of the SNP-effects on eGFRcrea and eGFRcys (i.e. age-effect in the interaction term) to be fairly linear when non-linear modelling of main age effect was applied (**Supplementary Figure 6 SA,B,C**). Of note, when the main age effect was modelled linearly, the SNP-effects on eGFRcrea and eGFRcys appeared to be non-linearly modified by age, which is a known problem in interaction analyses (data not shown); this supported the choice of the main age effect modelled non-linearly.

Supplementary Note S7: Narrow-sense heritability

We estimated SNP-based heritability (h^2) for eGFR-baseline and for eGFR-decline unadjusted and adjusted for eGFR-baseline using the genomic relatedness matrix restricted maximum likelihood (GREML) method as implemented in the GCTA software package (https://yanglab.westlake.edu.cn/software/gcta/#Overview). For this, we used individual participant data from UK Biobank for the ~15,000 unrelated individuals of European ancestry that had baseline and follow-up eGFR measurements available.

We estimated narrow-sense heritability (h^2) for eGFR-decline at 1% (standard error 2%, P = 0.31) and 5% for eGFR-decline adjusted for baseline (standard error 2.1%, P = 0.0075) and 20% (standard error 2.5%, P< 1.00×10^{-100}) for eGFR-baseline.

The small heritability for eGFR-decline in UK Biobank might derive from a large measurement error in eGFR-decline based on a study with only two measurements only 4 years apart. The larger heritability for eGFR-decline adjusted for eGFR-baseline compared to unadjusted for eGFR-baseline is reflective of the collider bias.

Supplementary Figure S1: Meta-analysis workflow. Shown is the meta-analysis workflow to capture the sequential recruitment and different suite of computed models (eGFR-decline unadjusted and adjusted for eGFR-baseline, "decline" and "decline adjusted"). In the first level, we conducted a meta-analysis of summary statistics across studies that were part of CKDGen since a long time ("old CKDGen studies", green boxes) and a meta-analysis across recently joined CKDGen studies ("new studies", blue boxes) and UK Biobank (orange box). In a second level, we meta-analyzed these two results. At each level, genomic-control (GC) correction was applied, when lambda was >1.00.
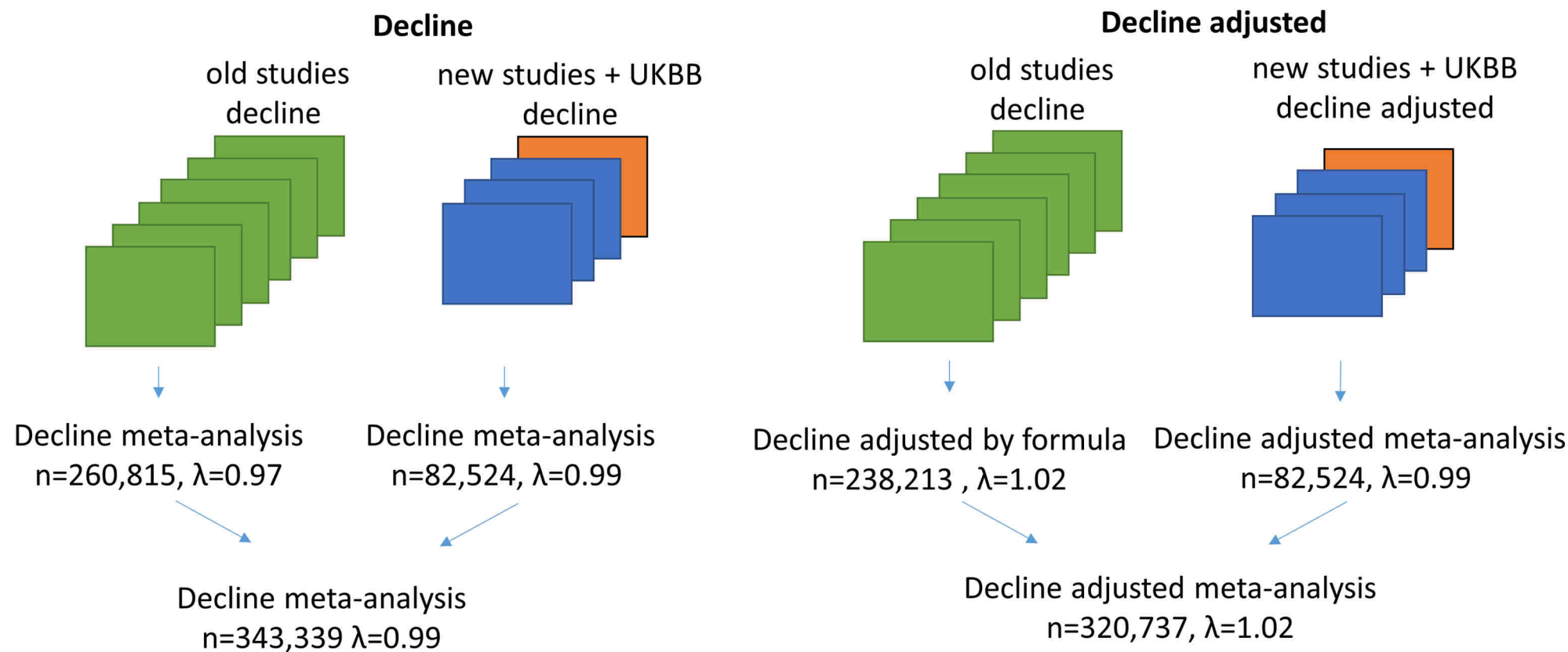
Supplementary Figure S2: Study-specific median annual eGFR-decline versus sample size, follow-up time and median age. Shown are, for each of the 62 studies, the study-specific median of annual eGFR-decline versus **(A)** number of individuals, **(B)** time to follow-up, and **(C)** median age at baseline. Whiskers represent interquartile range.
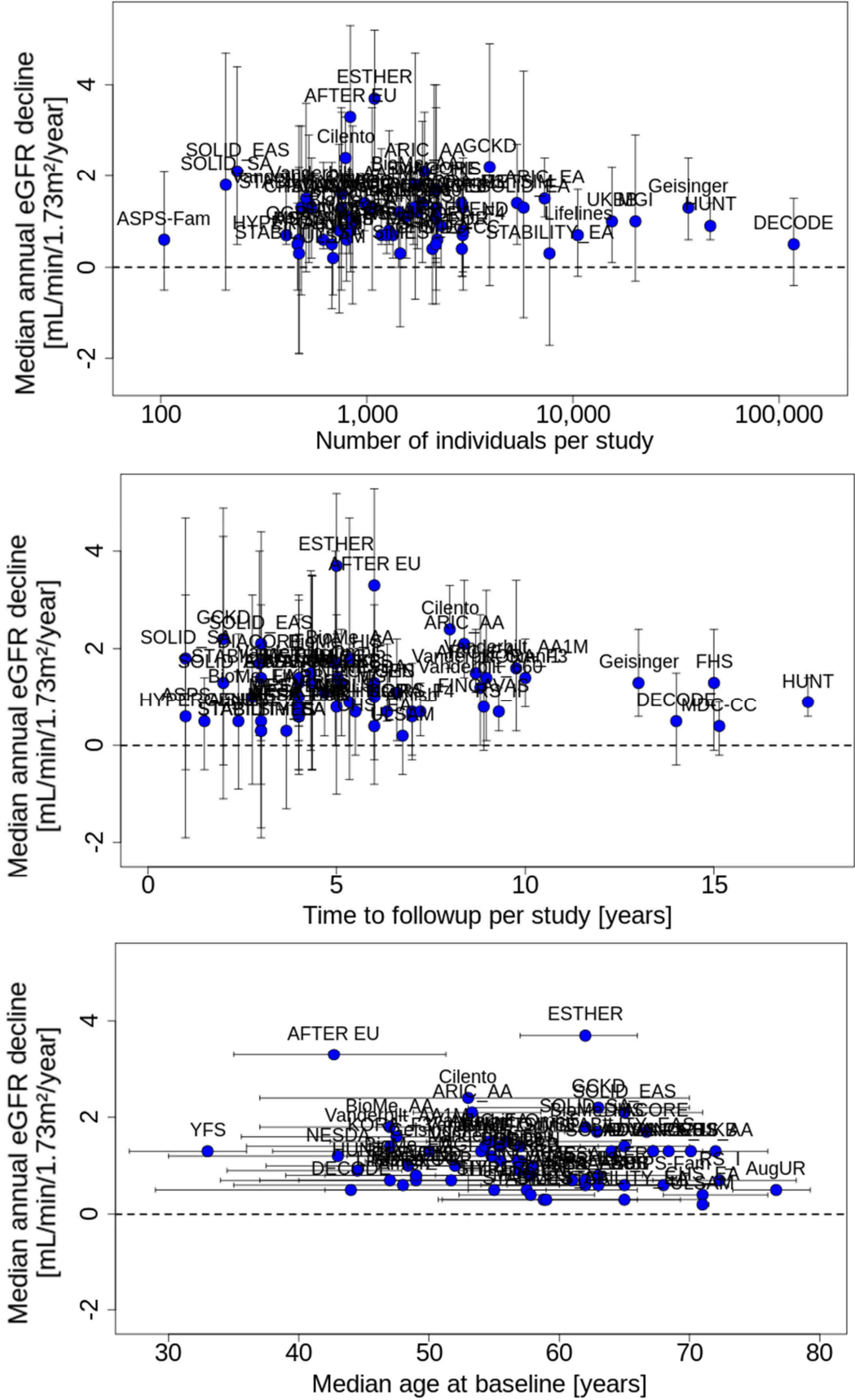
Supplementary Figure S3: No influence of alternative adjustments for age on eGFR-decline in UK Biobank. We explored alternative adjustments for age in UK Biobank (n=15,442, age range 40-70 years): **(A)** residuals of eGFR-decline adjusted for age, sex, and ln(eGFR-baseline) versus residuals of eGFR-decline adjusted for age, sex and residuals (ln(eGFR-baseline) adjusted for age and sex) and **(B)** residuals of eGFR-decline adjusted for age_centered (i.e. centered at 50 years) and sex with residuals of eGFR-decline adjusted for age_centered, (age_centered)^2 and sex. Alternative adjustments did not change the GWAS phenotype.
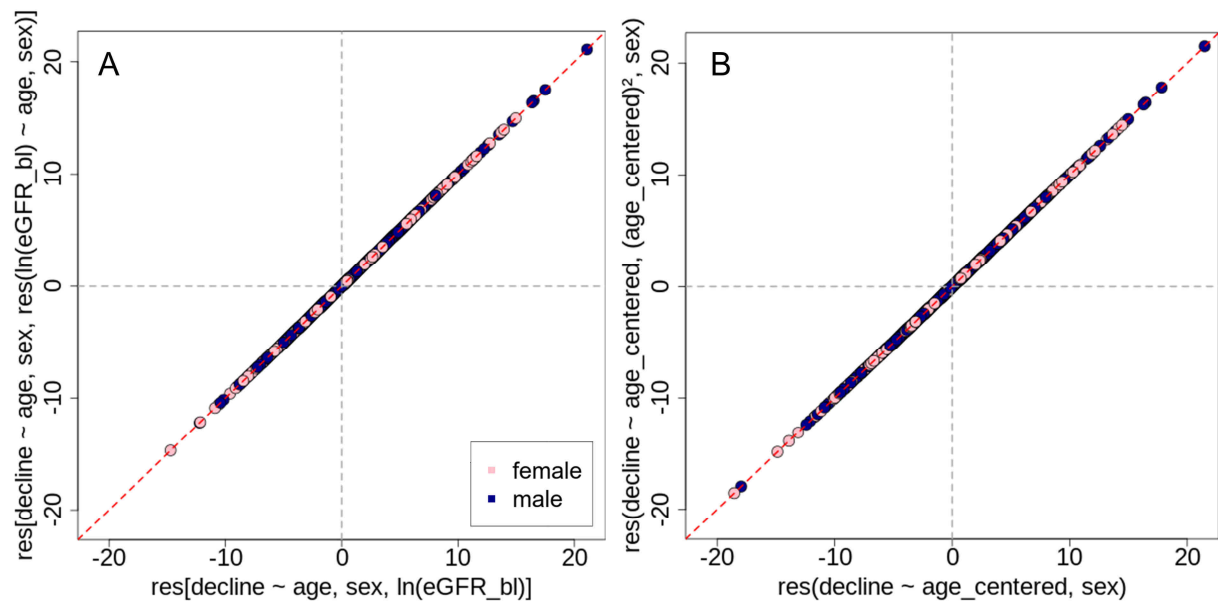
Supplementary Figure S4A: No influence from adjusting SNP-associations for eGFR-decline for diabetes mellitus (DM). We compared SNP-associations for eGFR-decline _with DM-adjustment_ with SNP-associations for eGFR-decline _without adjustment for DM_ in recently joined CKDGen studies, UK Biobank, several "old CKDGen studies", and their meta-analysis (total=103,970; **Supplementary Note S2**). Columns 1&2 show beta-estimates and standard errors (SE) among the 265 variants known for cross-sectional eGFR[S12], where we had a prior hypothesis that these might be associated with eGFR-decline. Columns 3&4 show betas and SEs genome-wide, where most SNP-associations are under the Null (i.e., not associated with eGFR-decline). Column 5 shows QQ-plots for P-values genome-wide. Coded allele is the cross-sectional eGFR-lowering allele, SNPs with minor allele frequency ≥0.05 are in green and with minor allele frequency <0.05 in orange. All SNPs have imputation quality>0.6 and MAC>10 for each study.
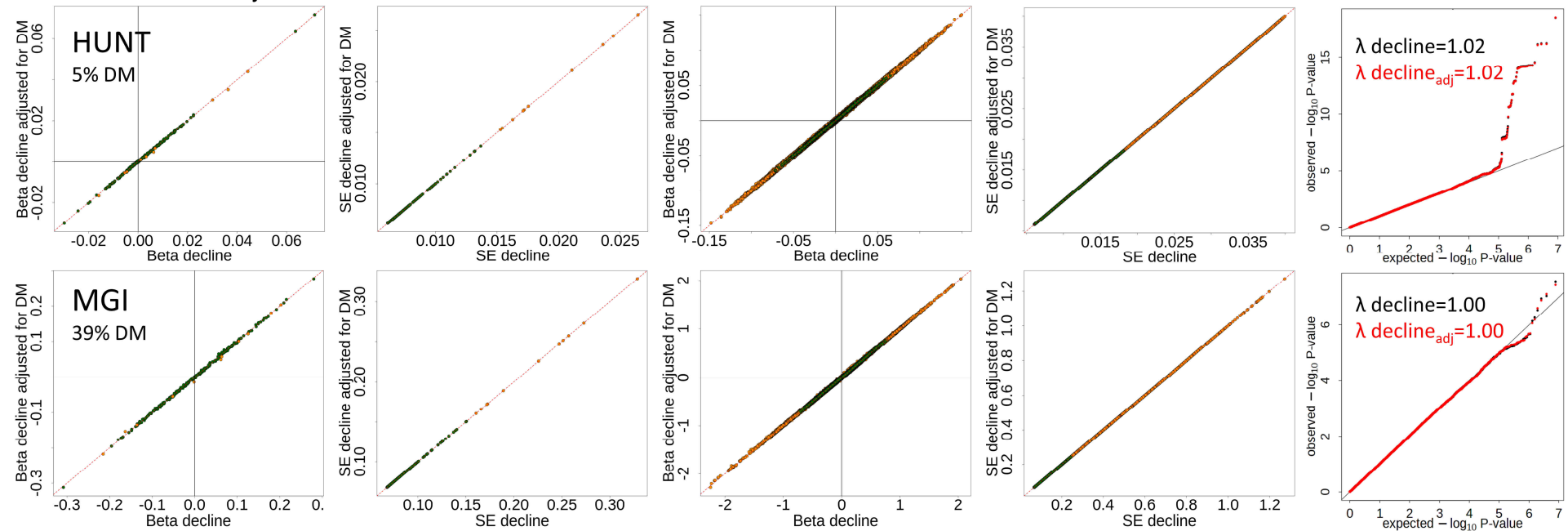
18

Supplementary Figure S4A: continued
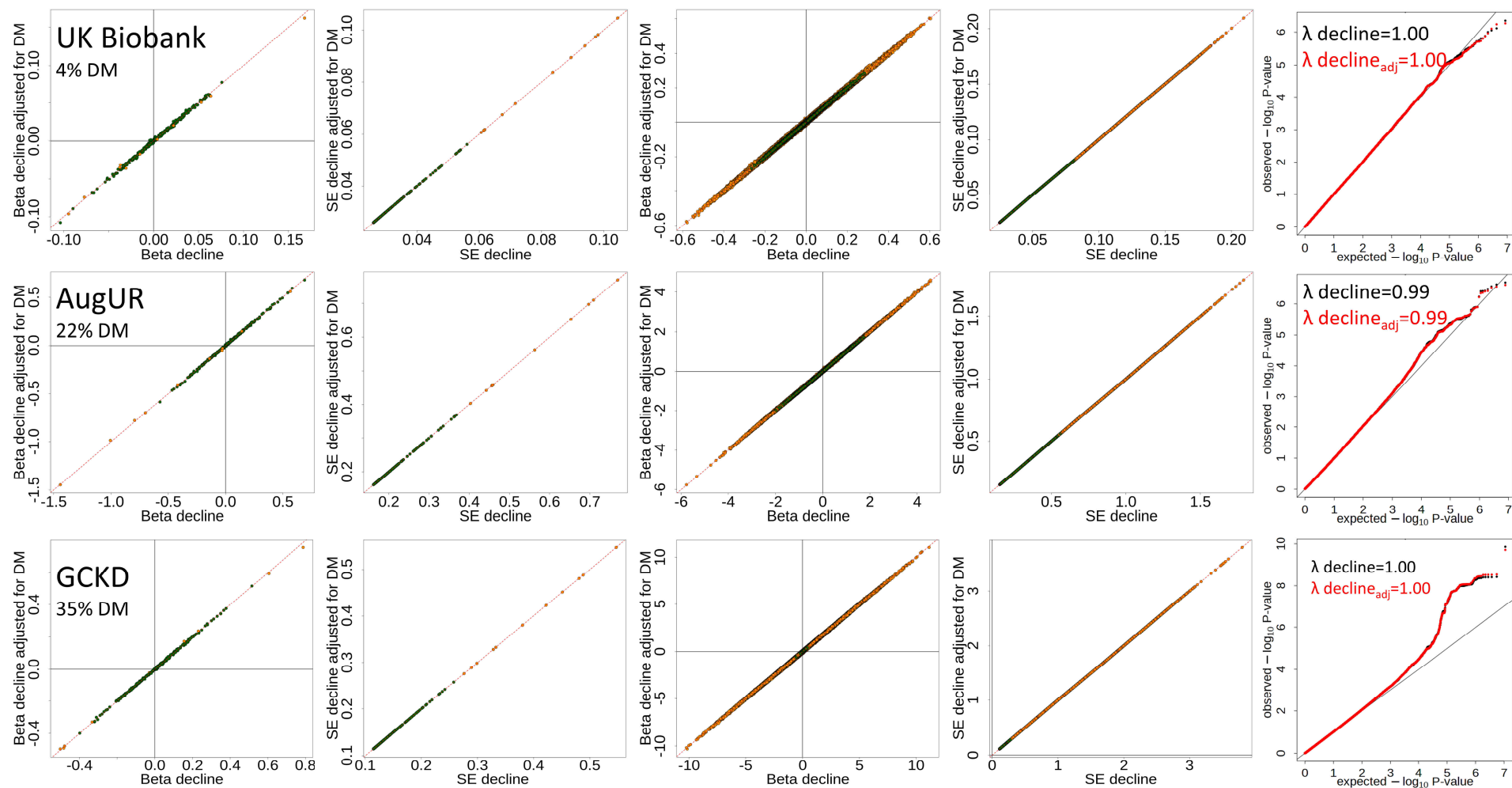
Supplementary Figure S4A: continued
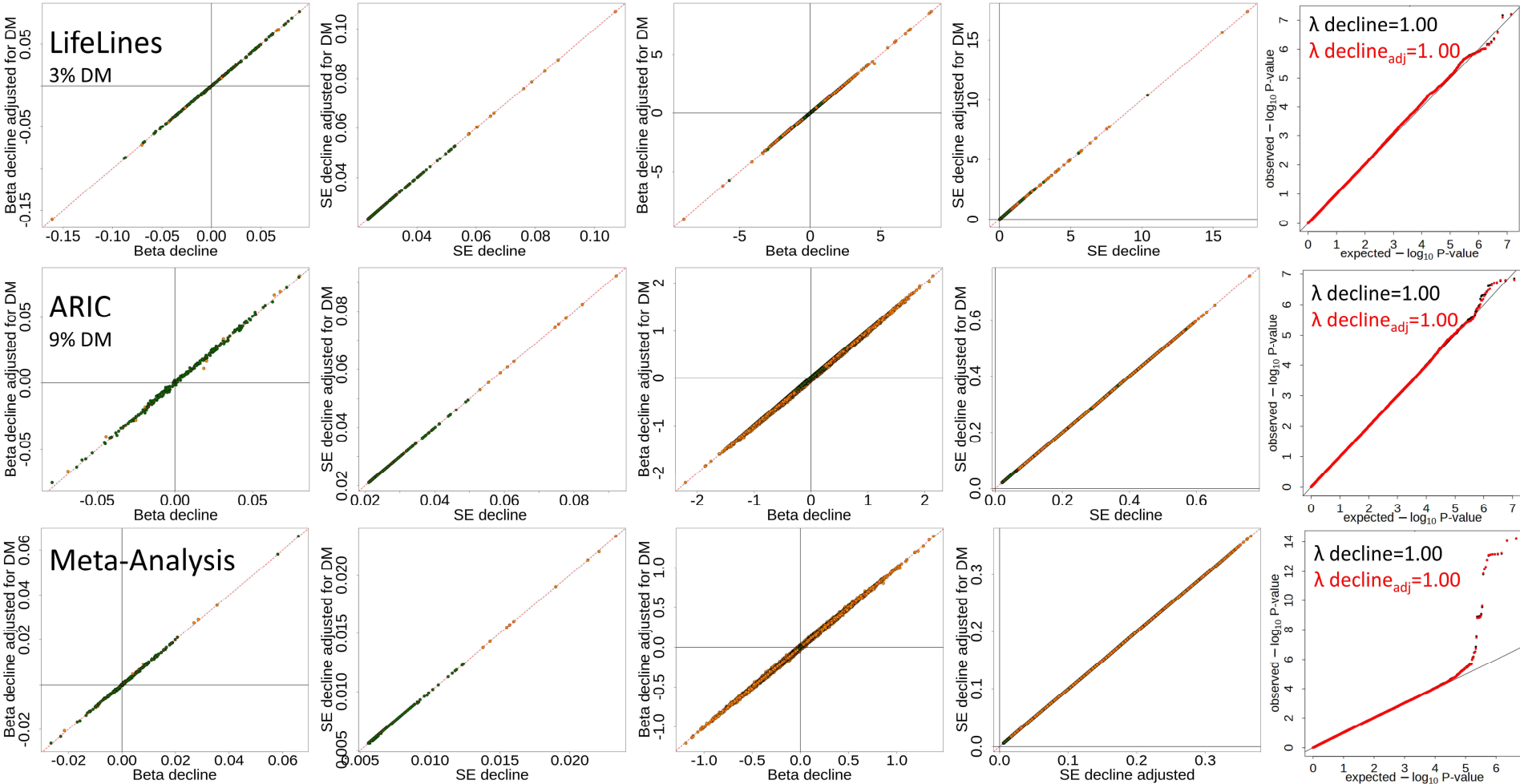
Supplementary Figure S4B: Differences between SNP-association for eGFR-decline unadjusted versus adjusted for eGFR-baseline We compared SNP-associations for eGFR-decline adjusted for eGFR-baseline with SNP-associations for eGFR-decline unadjusted for eGFR-baseline in recently joined studies, UK Biobank, several "old CKDGen studies", and their meta-analysis (total=103,970). Columns 1&2 show beta-estimates and standard errors (SE) among the 265 variants known for cross-sectional eGFR[S12], where we had a prior hypothesis that these might be associated with eGFR-decline. Columns 3&4 show betas and SEs genome-wide, where most SNP-associations are under the Null (i.e., not associated with eGFR-decline). Column 5 shows QQ-plots for P-values genome-wide. Coded allele is the cross-sectional eGFR- lowering allele, SNPs with minor allele frequency ≥0.05 are in green and with minor allele frequency <0.05 in orange. All SNPs have imputation quality>0.6 and MAC>10 for all studies.

Supplementary Figure S4B: continued

Supplementary Figure S4B: continued

Supplementary Figure S4C: Validation of formula-derived adjustment for eGFR-baseline in eGFR-decline associations (part 1). We compared SNP-associations for eGFR-decline adjusted for eGFR-baseline by model with SNP-associations for eGFR-decline adjusted for eGFR-baseline by formula (using beta-estimates for eGFR-baseline) in recently joined studies, UK Biobank, several "old CKDGen studies", and their meta-analysis (total=103,970). Columns 1&2 show beta-estimates and standard errors (SE) among the 265 variants known for cross-sectional eGFR[S12], where we had a prior hypothesis that these might be associated with eGFR-decline. Columns 3&4 show betas and SEs genome-wide, where most SNP-associations are under the Null (i.e., not associated with eGFR-decline). Column 5 shows QQ-plots for P-values genome-wide. Coded allele is the cross-sectional eGFR-lowering allele, SNPs with minor allele frequency ≥0.05 are in green and with minor allele frequency <0.05 in orange. All SNPs have imputation quality>0.6 and MAC>10 for all studies.

Supplementary Figure S4C: continued
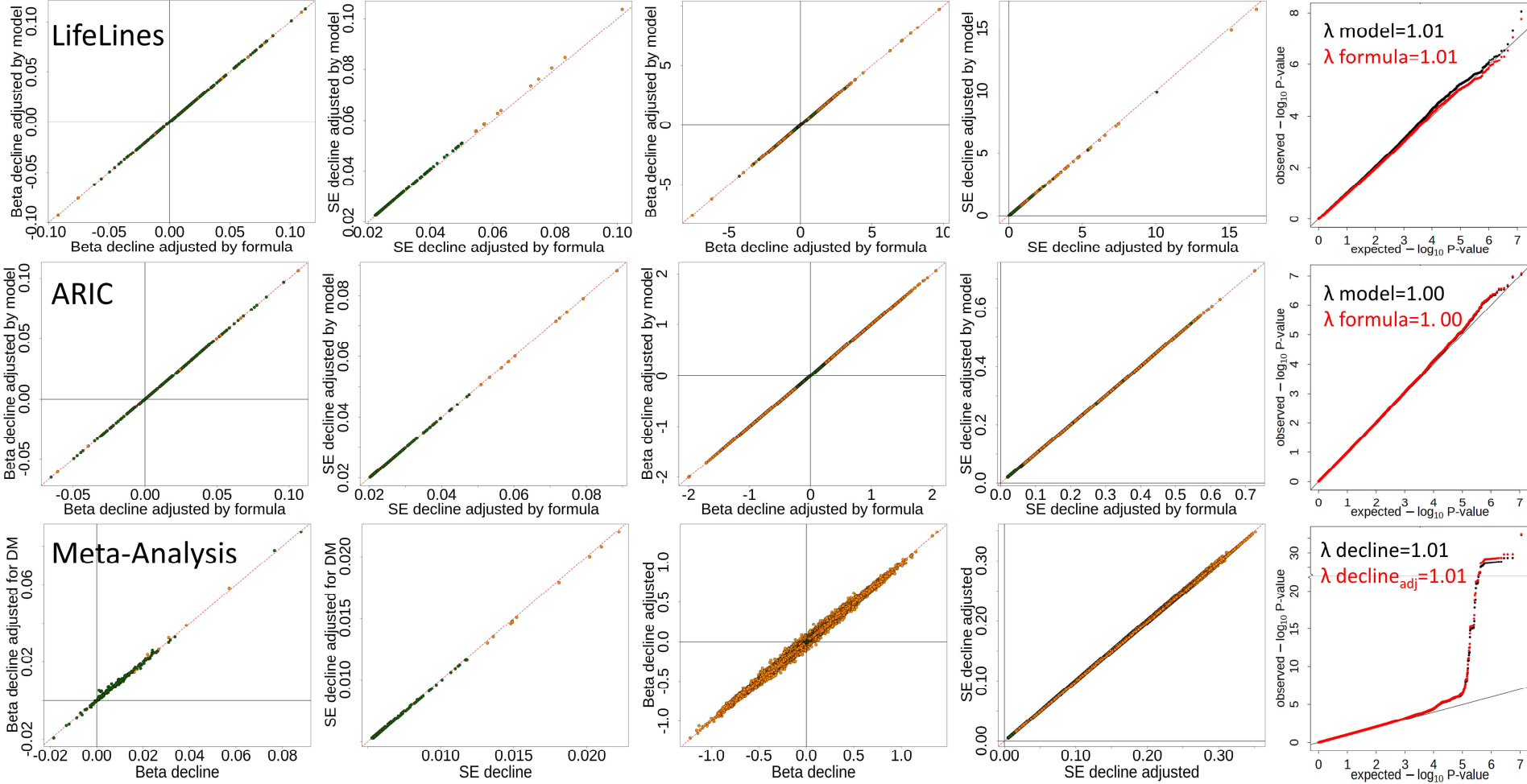
Supplementary Figure S4D: Validation of formula-derived adjustment for eGFR-baseline in eGFR-decline associations (part 2). In "old CKDGen studies", sample sizes were typically larger for cross-sectional eGFR than for baseline eGFR (i.e. restricted to individuals in follow-up). We compared SNP-associations for eGFR-decline <u>adjusted for eGFR-baseline by model</u> with SNP-associations for eGFR-decline <u>adjusted for eGFR-baseline by formula</u> using beta-estimates for <u>cross-sectional eGFR</u> in six "old" CKDGen studies. Columns 1&2 show beta-estimates and standard errors (SE) among the 265 variants known for cross-sectional eGFR[S12], where we had a prior hypothesis that these might be associated with eGFR-decline. Columns 3&4 show betas and SEs genome-wide, where most SNP-associations are under the Null (i.e., not associated with eGFR-decline). Column 5 shows QQ-plots for P-values genome-wide. Coded allele is the cross-sectional eGFR-lowering allele, SNPs with minor allele frequency ≥0.05 are in green and with minor allele frequency <0.05 in orange. All SNPs have imputation quality>0.6 and MAC>10 for all studies.
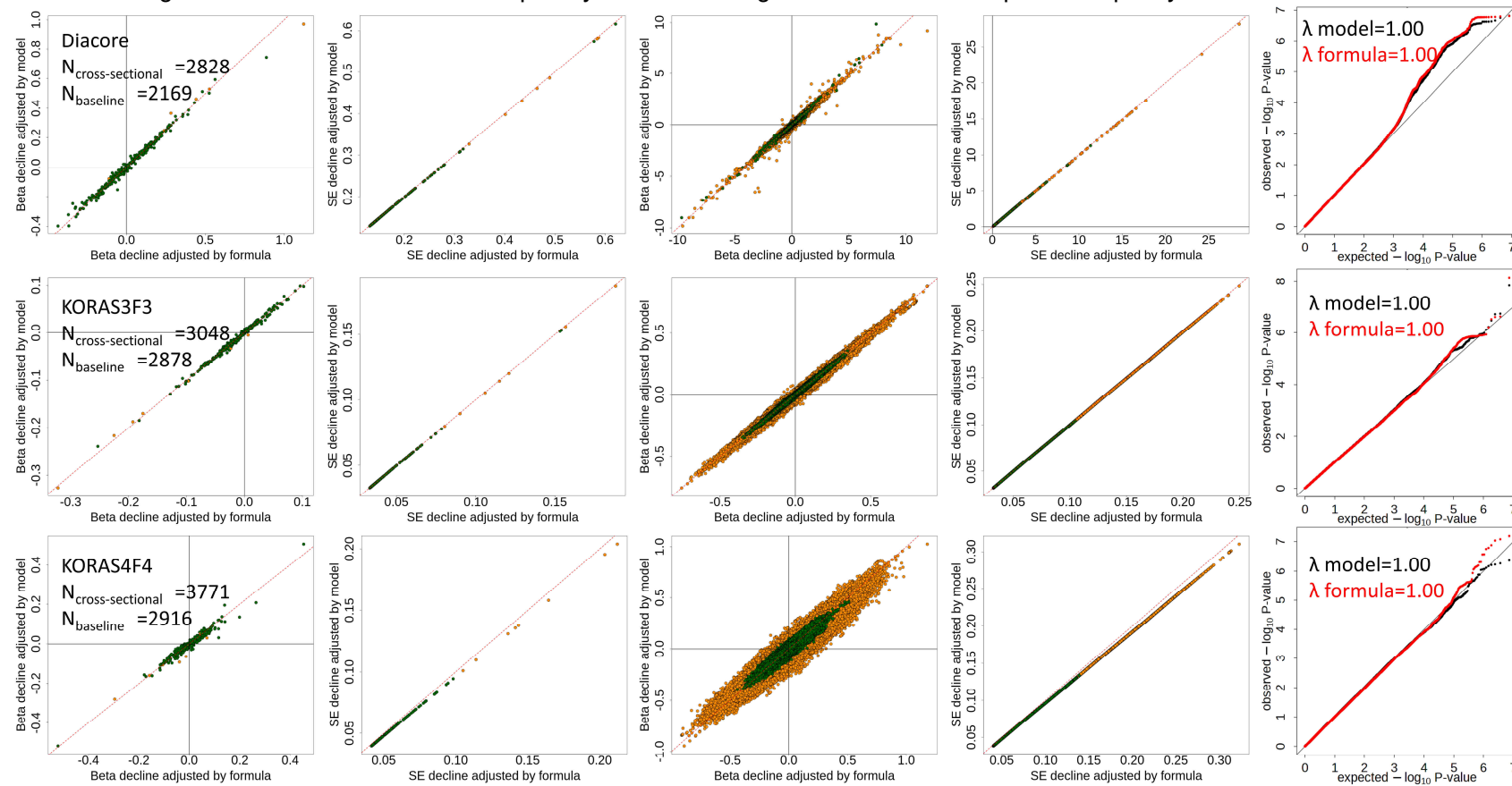
Supplementary Figure S4D: continued
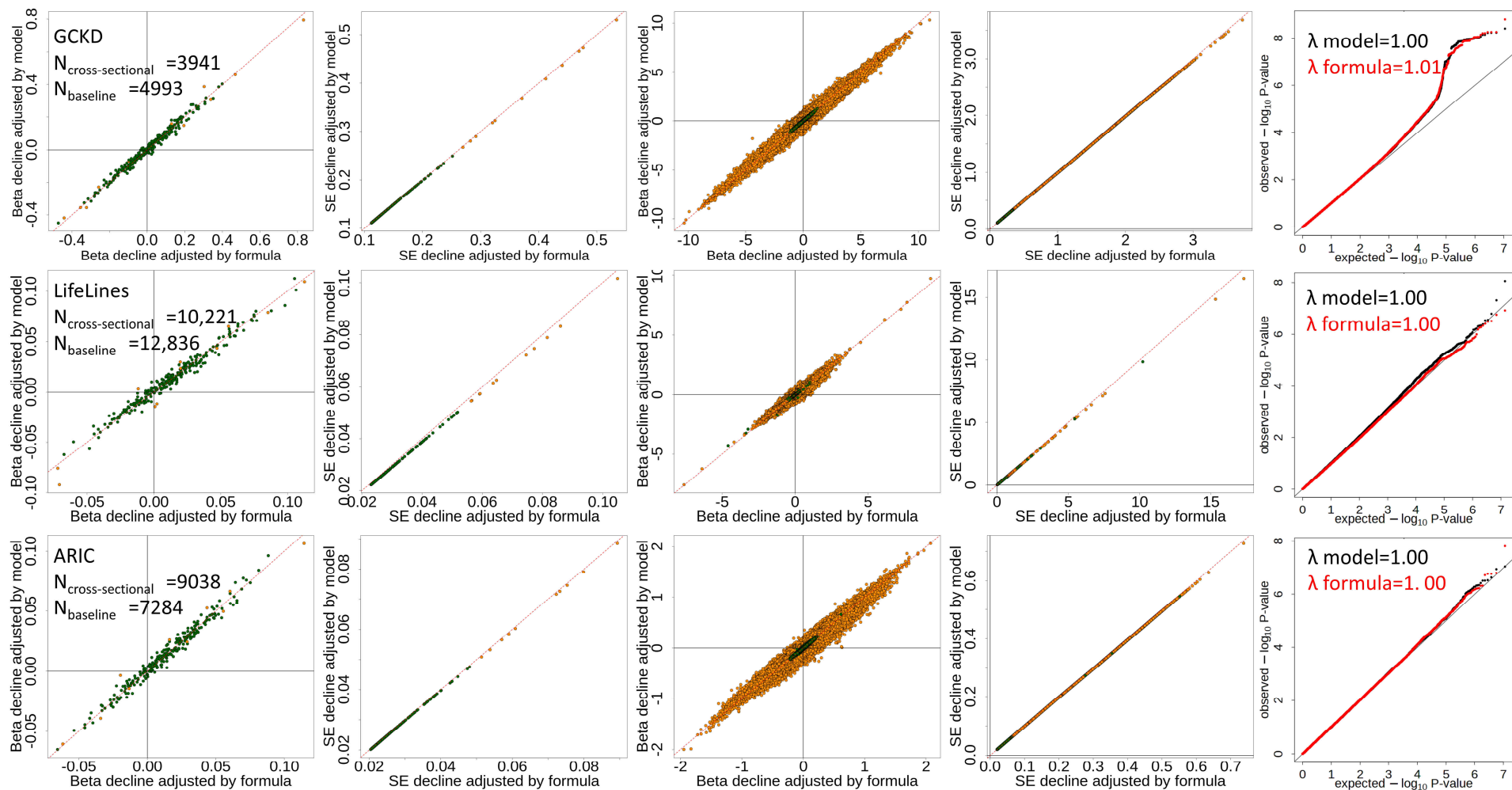
Supplementary Figure S5A: Region plots of the 4 variants in 3 loci identified for eGFR-decline unadjusted for eGFR-baseline. Shown are regional association plots (1st column) for cross-sectional eGFR[S12] ("eGFRcrea", n up to 765,348), (2nd and 3rd column) for eGFR-decline unadjusted for eGFR-baseline ("decline"; n up to 343,339; blue dashed line P=0.05/263=1.90x10[-4] in 2nd column and P=0.05 in 3rd column), and (4th column) for eGFR-decline adjusted for eGFR-baseline ("decline$_{adj}$"; n up to 320,737). Reference variants are the cross-sectional eGFR lead variant (1st and 2nd column) and the decline$_{adj}$ lead variant (i.e. variant with the smallest P-value for decline$_{adj}$; 3rd and 4th column). Red lines indicate P=5.00x10[-8]. The decline signals coincide with the cross-sectional eGFR signals; decline lead variants are the same or highly correlated with cross-sectional eGFR lead variants (r²=same, same, 1.00 and 0.93 for *UMOD-PDILT* (2), *PRKAG2* and *SPATA7*, respectively).
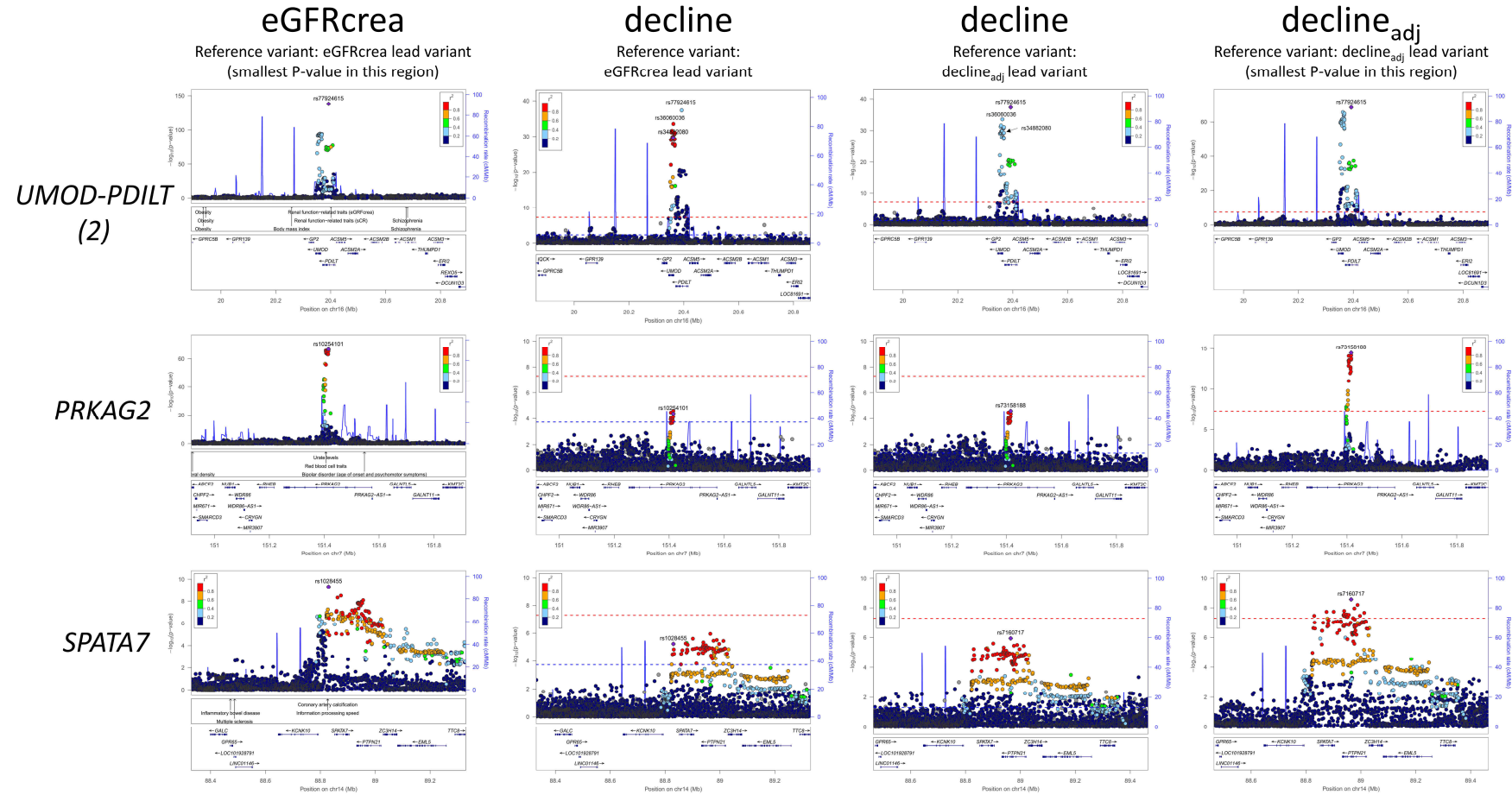
Supplementary Figure S5B: Regions of the 5 variants in 5 loci identified from GWAS for eGFR-decline adjusted for eGFR-baseline with significant association for eGFR-decline unadjusted for eGFR-baseline. Shown are regional association plots (1st column) for cross-sectional eGFR[S12] ("eGFRcrea", n up to 765,348), (2nd and 3rd column) for eGFR-decline unadjusted for eGFR-baseline ("decline"; n up to 343,339; blue dashed line $P=0.05/263=1.90\times10^{-4}$ in 2nd column and $P=0.05$ in 3rd column), and (4th column) for eGFR-decline adjusted for eGFR-baseline ("decline$_{adj}$"; n up to 320,737). Highlighted are lead variants for cross-sectional eGFR[S12] (1st and 2nd column; for *C15ORF54*, using the 2nd signal lead variant) or the decline$_{adj}$ lead variant (3rd and 4th column). Red lines indicate $P=5.00\times10^{-8}$. Signals for decline$_{adj}$ coincide with signals for cross-sectional eGFR.
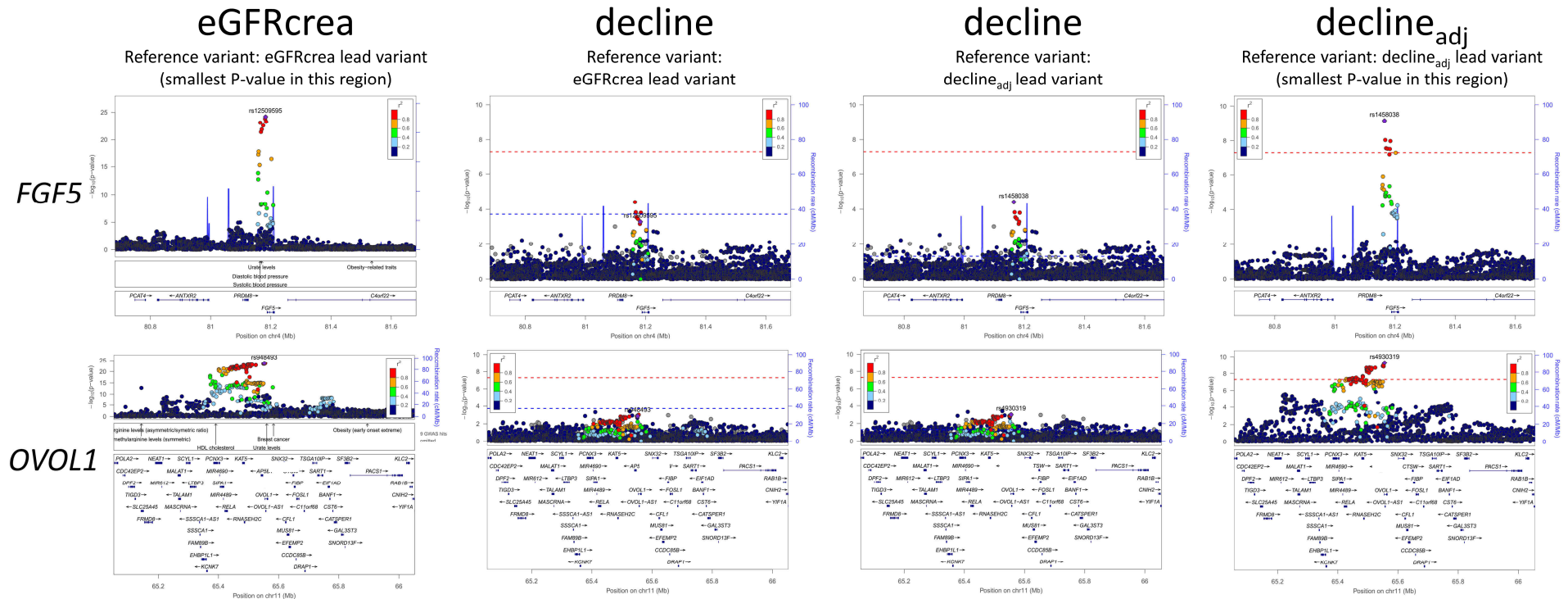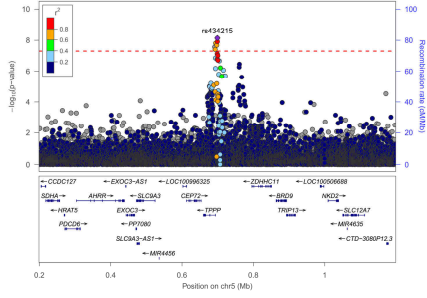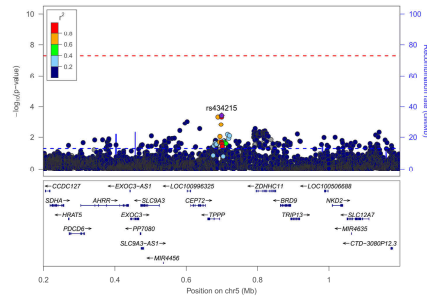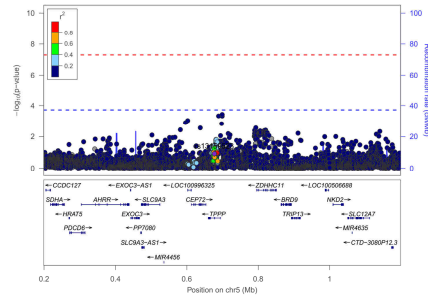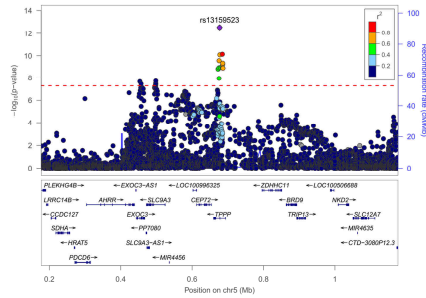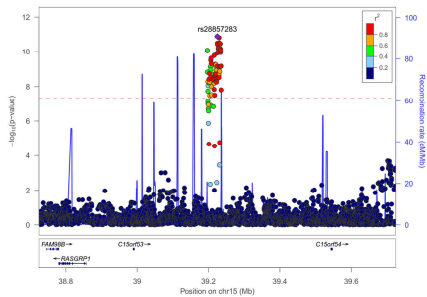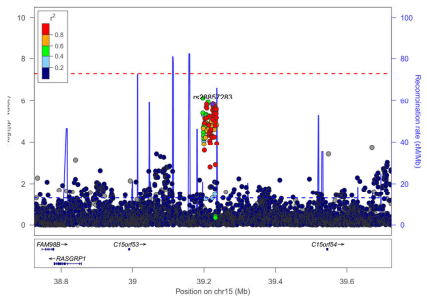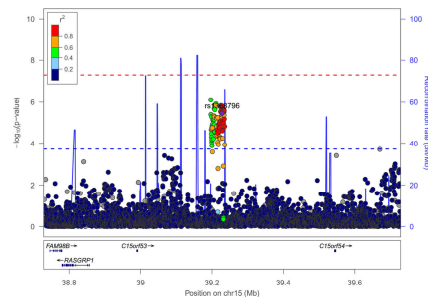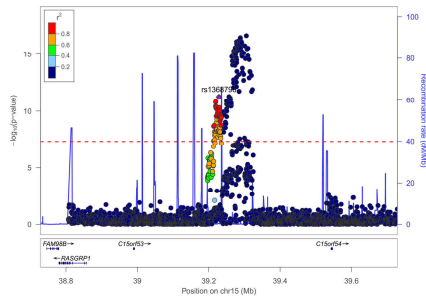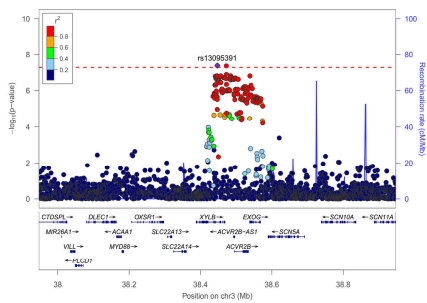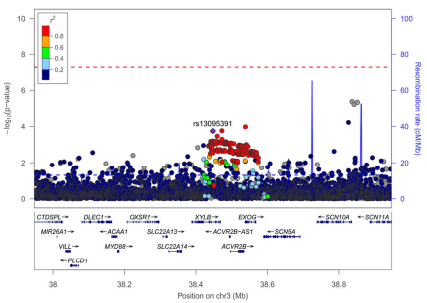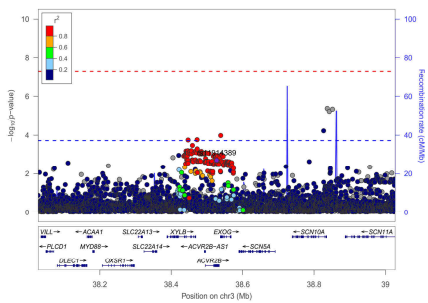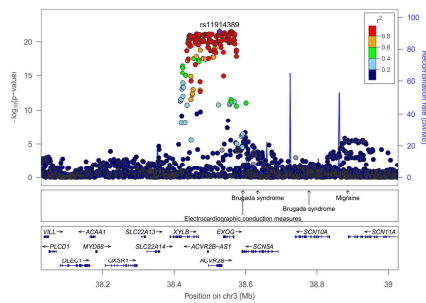
Supplementary Figure S5C: Regions of the 3 variants in 3 loci identified from GWAS for eGFR-decline adjusted for eGFR-baseline without significant association for eGFR-decline unadjusted for eGFR-baseline. Shown are regional association plots (1st column) for cross-sectional eGFR[S12] ("eGFRcrea", n up to 765,348), (2nd and 3rd column) for eGFR-decline unadjusted for eGFR-baseline ("decline"; n up to 343,339; blue dashed line P=0.05/263=1.90x10^{-4} in 2nd column and P=0.05 in 3rd column), and (4th column) for eGFR-decline adjusted for eGFR-baseline ("decline$_{adj}$"; n up to 320,737). Highlighted are lead variants for cross-sectional eGFR[S12] (1st and 2 column) and decline$_{adj}$ lead variants (3rd and 4th column). Red lines indicate P=5.00x10^{-8}. Signals for decline$_{adj}$ coincide with signals for cross-sectional eGFR; there is no association for decline (unadjusted) in these regions.
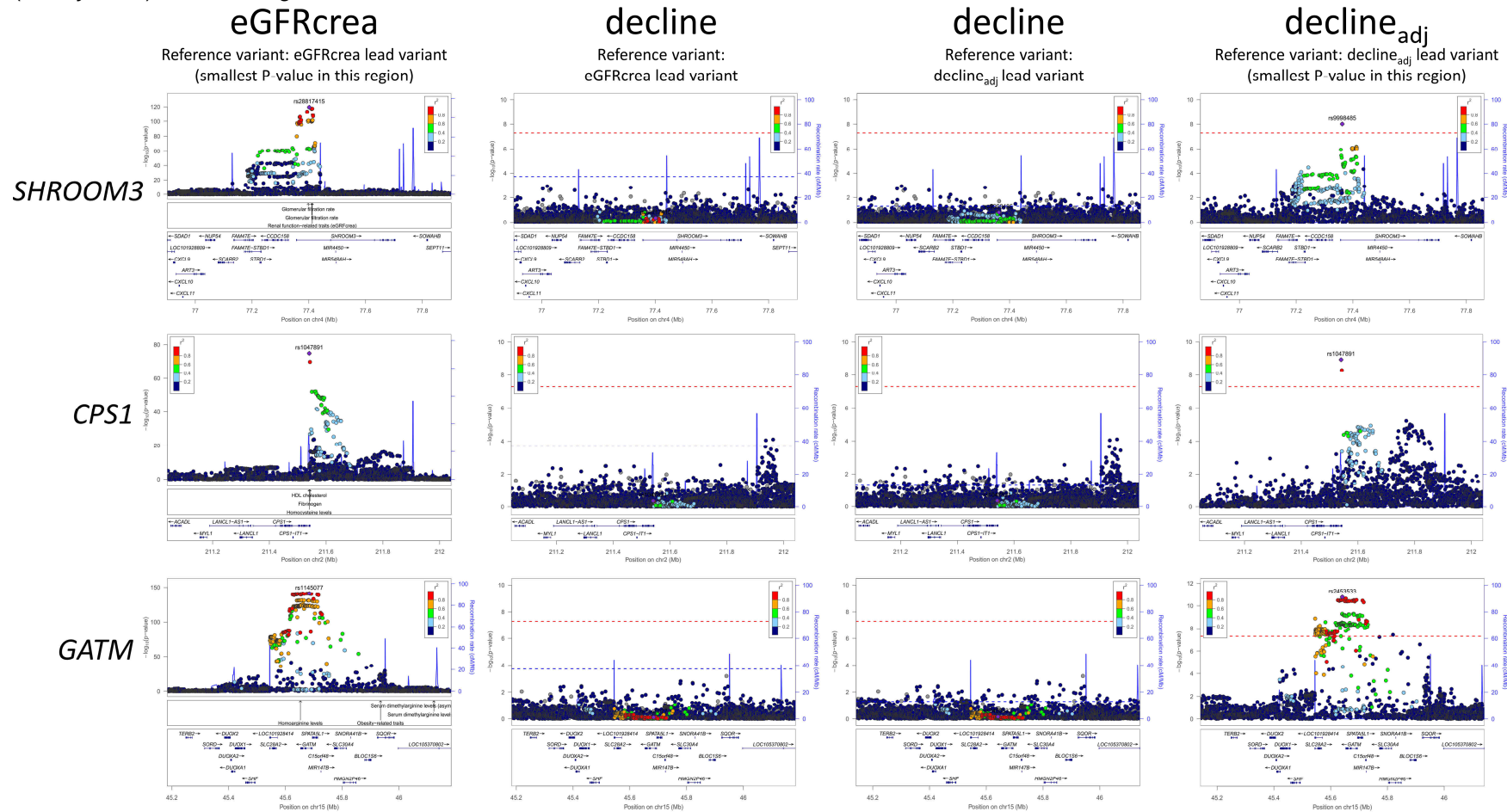
Supplementary Figure S6: Age-dependency of cross-sectional eGFR and age-dependency of SNP-effects on cross-sectional eGFR in UK Biobank. We conducted SNP-by-age interaction analyses on cross-sectional eGFRcrea and eGFRcys in individuals from UK Biobank that were independent from the GWAS (n=351,462; i.e. excluding the 15,442 individuals in the eGFR-decline GWAS) using linear regression with covariates sex, age, SNP, SNP-by-age and outcome eGFRcrea or eGFRcys. The SNP-effect was modelled as linear dosage effect (for main effect and in interaction term; i.e. additive genetic effect per allele). Age was centered at 50 years and modelled linearly as well as allowing for a smooth non-linear change by age. For cross-sectional eGFRcrea (1st row) and eGFRcys (2nd row), we show the age-dependency **(Z)** of the main age effect on eGFRcrea and eGFRcys, **(A)** on the SNP-effects of the 4 variants identified for eGFR-decline (unadjusted for eGFR-baseline), **(B)** on the SNP-effects of the 5 variants identified for eGFR-decline adjusted for eGFR-baseline with significant association for eGFR-decline unadjusted for eGFR-baseline, and **(C)** on the SNP-effects of the 3 variants identified for eGFR-decline adjusted for eGFR-baseline without significant association for eGFR-decline unadjusted for eGFR-baseline. In **A-C**, the main age effect was modelled non-linearly (to avoid residual confounding) and the interaction effects modelling the age-dependency of the SNP-effect linearly (green lines) are the ones reported in **Table 3**.
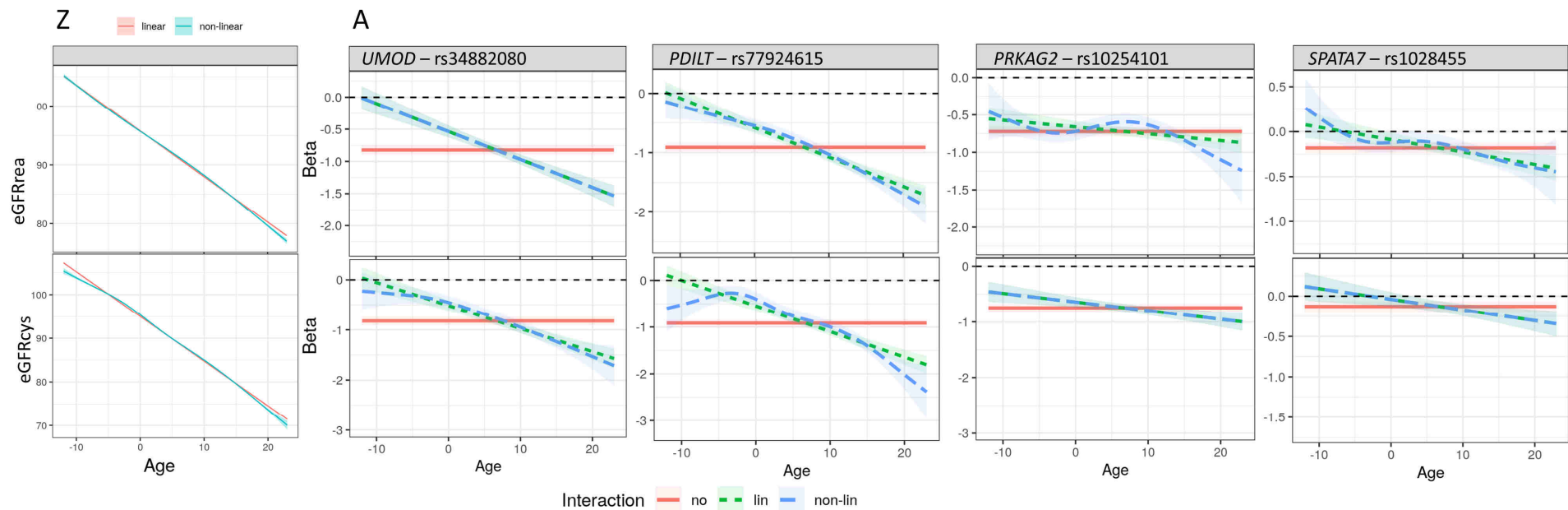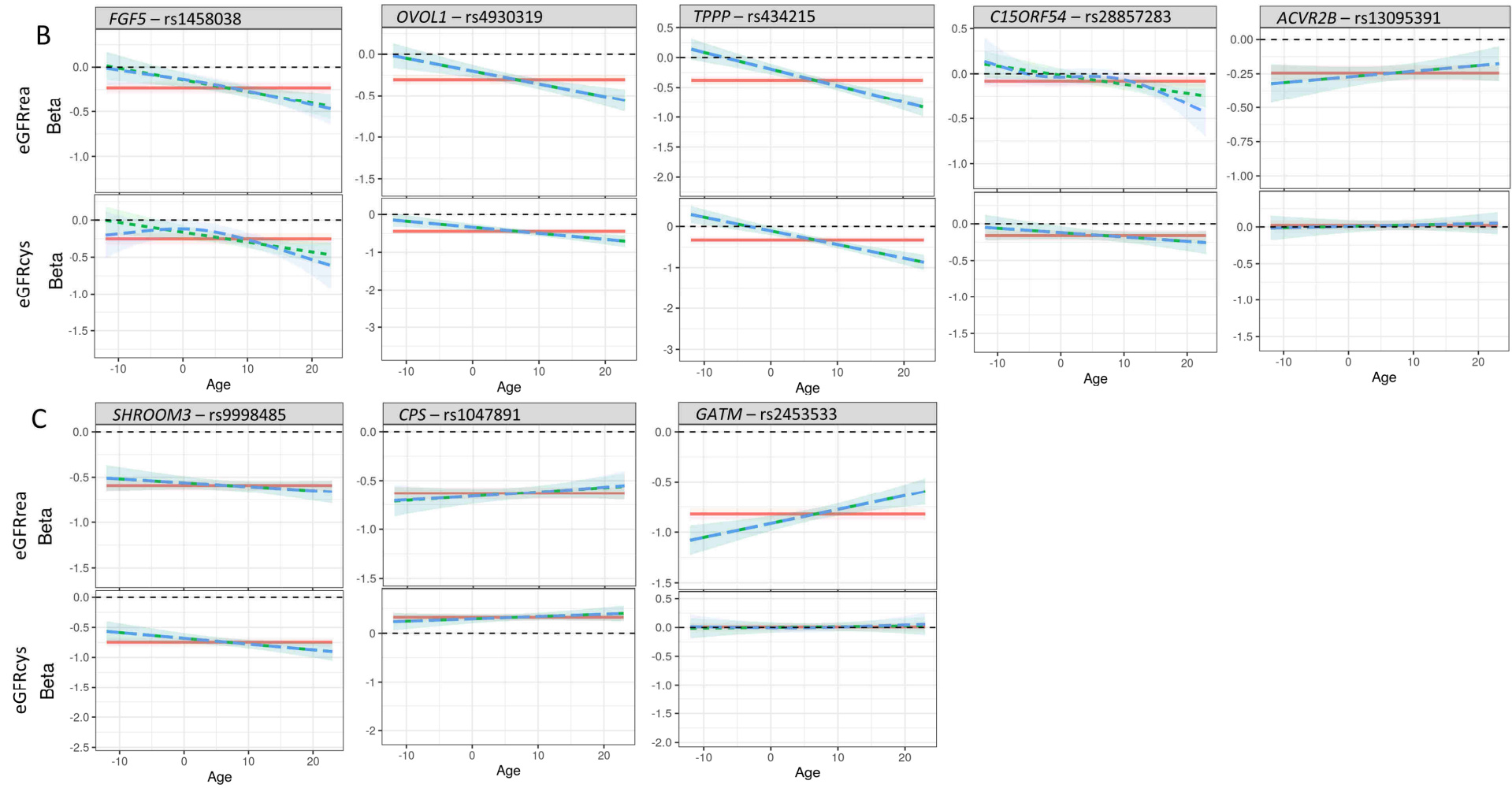
Supplementary Table S1: Description of participating studies: study design						
Study	Full name of the study	Subgroup	Ancestry (EA/AA/HIS/EAS/SA)	Study Design (if not population-based, please specify selection and/or enrichment strategy)	Important study references, e.g. design paper (PMID)	Serum creatinine assay and year of measurement, baseline
ADVANCE	Action in Diabetes and Vascular disease: preterAx and diamicroN mr Controlled Evaluation	5	EA	factorial, multicentre, randomised controlled trial, with a 5- to 6-year follow-up.	11848259	Jaffe, 2001-2003
		6	EA			
		UKB	EA			
AFTER EU	AFTER (EURAGEDIC) European Rational Approach for the Genetics of Diabetic Complications		EA	Adult onset Type 1 Diabetes	18496510, 20357380	Modified Jaffe
Amish	Amish Studies		EA	Population based "founder" cohort	18440328, 26374108, 15621217	Modified kinetic Jaffe reaction
ARIC	Atherosclerosis Risk in Communities study	AA	AA	Population-based	2646917	Modified kinetic Jaffé reaction, 1989
		EA	EA			
ASPS	Austrian Stroke Prevention Study		EA	Population-based	10408549, 7800110	Modified kinetic Jaffe reaction, 1991 - 2005
ASPS-Fam	Autrian Stroke Prevention Family Study		EA	Family-based	25309438, 25443291	Modified kinetic Jaffe reaction, 2006 - 2012
BioMe	BioMe™ BioBank Program	Omni AA	AA	Population-based	25349204	Jaffe, 2008
		Omni EA	AA			
		Omni HA	HIS			
CHS	Cardiovascular Health Study	AA	AA	Population-based	1669507	Colorimetric method on a Kodak Ektachem 700 Analyzer (Eastman Kodak, Rochester, NY), 1989-90 and 1992-93
		EA	EA			
Cilento	Cilento Study		EA	Population-based, Isolated Population Study	17476112, 19550436	Jaffe, 2005
DECODE	deCODE genetics/Amgen		EA	Population-based	20686651, 25082825	Ensymatic and modified kinetic Jaffe reaction assay since 1997
DIACORE	DIAbetes COhoRtE		EA	Prospective cohort study of patients with diabetes mellitus type 2	23409726	Serum Creatinine was measured 2010-2013 using an enzymatic assay traceable to NIST.
ESTHER	Epidemiological investigation of the chances of preventing, recognizing early and optimally treating chronic diseases in an elderly population		EA	Population-based	23446902, 15578318	Kinetic Jaffe-method, 2000 - 2002
FHS	The Framingham Heart Study		EA	Community- and family-based	5921755, 1208363, 17372189	Modified Jaffe method
FINCAVAS	The Finnish Cardiovascular Study		EA	Fincavas follow-up cohort of consecutive patients undergoing exercise stress test	16515696	Entzymatic photometric, 1992-2015
GCKD	German Chronic Kidney Disease study		EA	Included are European ancestry CKD patients aged 18-74 years with an eGFR between 30–60 mL/min per 1.73 m2 or an eGFR >60 mL/ min per 1.73 m2 and a urinary albumin-to-creatinine ratio (UACR) >300 mg/g, albuminuria >300 mg/day, a urinary protein-to-creatinine ratio >500 mg/g, or proteinuria >500 mg/day	21862458, 25271006	Serum creatinine was measured using the Ceratinine plus enzymatic assay (Roche) on a Modular (P) analyzer in 2012
Geisinger Research (MyCode)	MyCode Community Health Initiative		EA	Population-based	26866580	Enzymatic method done by Roche Cobas instruments, 1996+
HANDLS	Healthy Aging in Neighborhoods of Diversity across the Life Span study		AA	Population-based prospective longitudinal study	20828101	Modified Jaffe 2004-2009
HYPERGENES	Hypergenes - European Network for Genetic-Epidemiological Studies	controls	EA	Case-control for Hypertension	22184326	Jaffe assay 2002
Jackson Heart Study (JHS)	Jackson Heart Study		AA	Community and family-based	16320381	IDMS calibrated serum creatinine was used from visit 1 and visit 3… creatinine measurements were made from 2000 on but calibration to the same standard was done in 2015 (see PMID: 25806862 for a full description).
JMICC	Japan Multi-institutional Collaborative Cohort (J-MICC) Study		EAS	Population-based	17696755, 32963210	Enzymatic method, 2007-2010
KORA	Cooperative Health Research in the Augsburg Region	F3	EA	Population-based	16032514	Modified kinetic Jaffe reaction, 1994
		F4	EA			
Lifelines	Lifelines Cohort Study		EA	Population-based	18075776, 25502107, 26333164	Enzymatic, IDMS traceable, Roche (Modular); 2006-2013
MDC-CC	Malmö Diet and Cancer Study-Cardiovascular Cohort		EA	Population-based	11916347	Jaffé method and the IDMS-traceable standard was used
MESA	Multi-Ethnic Study of Atherosclerosis	AFR	AA	Population-based without CVD	12397006	Baseline is year 2002, exam 2 2004, exam 3 2005 and exam 4 2007. All assays rate reclectance spectrophotometry using thin film adaptation of the creatine aminohydrolase method on the Vitros analyzer (Johnson and Johnson Clinical Diagnositcs)
		EAS	EAS			
		EUR	EA			
		HIS	HIS			
METSIM	Metabolic Syndrome in Men study		EA	Population-based	28119442	Kinetic Jaffé method, 2005-2010
NESDA	Netherlands Study of Depression and Anxiety		EA	Population-based, predominantly cases with major depression	18763692	Partly Jaffe, partly enzymatic; 2004-2007
OGP	Ogliastra Genetic Park Study		EA	Population-based	20823129	Colorimetric method Jaffè without deproteinization (Biotecnica instruments).Creatinine forms a colored orange-red complex in an alkaline picrate solution. The difference in absorbance at fixed times during conversion is proportional to the concentration of creatinine in the sample. 2005-2008
PIVUS	Prospective Investigation of Vasculature in Uppsala Seniors		EA	Population-based	16141402	Kinetic jaffe method
POPGEN	POPGEN control sample		EA	Population-based	16490960	Serum creatinine was measured 2005-2008 using an enzymatic assay
PREVEND	Prevention of Renal and Vascular End-stage Disease study		EA	Population-based	12356629	An isotope dilution mass spectrometry (IDMS) traceable enzymatic method on a Roche Modular analyzer using reagents and calibra- tors from Roche (Roche Diagnostics, Mannheim, Germany) '97-'98
RS	Rotterdam Study	I	EA	Population-based	29064009	Enzymatic assay, 1999
		II	EA			Enzymatic assay, 2000
		III	EA			Enzymatic assay, 2006
SHIP	Study of Health in Pomerania	1	EA	Population-based	20167617	Jaffe, 2002
SiMES	Singapore Malay Eye Study		EAS	Population-based	17365815, 21490949	Jaffe, 2004-2007
SINDI	Singapore Indian Eye Study		EAS	Population-based	19995197, 24244560	Jaffe, 2007-2009
SOLID-TIMI 52	SOLID-TIMI 52	EA	EA	Clinical trial	21982651	Jaffe, 2010
		EAS	EAS			
		SA	SA			

STABILITY	STabilization of Atherosclerotic plaque By Initiation of darapLadIb TherapY	EA	EA	Clinical trial	24678955, 20934559	Jaffe, 2009
		EAS	EAS			
		SA	SA			
ULSAM	Uppsala study of adult men		EA	Population-based	21335440	Kinetic jaffe method
Vanderbilt	Vanderbilt BioVU	660	EA	Population-based with enrichment for a variety of disease studies	18500243	Extracted from clinical records
		AA1M	AA			
		Omni1	EA			
		Omni5	EA			
YFS	The Young Finns Study		EA	Population-based	18263651, 23069987	Serum creatinine was determined spectrophotometrically by the Jaffé method (picric acid; Olympus Diagnostica GmbH) from frozen plasma samples. Year 2001.
AugUR	The German AugUR study		EA	Prospective cohort study in the elderly	26489512	Serum Creatinine was determined on a encymatic Siemens-Kit ECREA, 2018
HUNT	Trøndelag Health Study, Norway		EA	Population-based	22879362	Modified kinetic Jaffé reaction, 1995-1997
MGI	Michigan Genomics Initiative		EA	Hospital-based		Jaffe, variable year of measurement
UKBB	Uk Biobank		EA	Population-based	25826379	Enzymatic analysis on a beckman Coulter AU5800

AA: African American ancestry; EA: European ancestry; HIS: Hispanics; SA: South Asian ancestry; EAS: East Asian ancestry

Supplementary Table S2: Description of participating studies: genotyping and imputation												
Study	Exclusions prior to genotyping and/or ge	Genotyping Array	Genotype calling	QC filters for genotyped SNPs used for imputation	No of SNPs used for imputation	Pre-phasing software[1]	Imputation	Imputation reference panel	Filtering of imputed genotypes	Software used for GWAS[3]	Handling of population stratification	Type of reported imputation quality
ADVANCE	Ethnic outliers, sex mismatches, call rate < 95%	Affymetrix 5.0, Affymetrix 6.0, Affymetrix UKB	Affymetrix power tools 1.17.0	avg_het <23% or >30%; call rate <97%; MAF <1%; snp call rate <95%; HWE <0.001;	Affymetrix 5.0 : 363,062; Affymetrix 6.0 : 702,628; Affymetrix UKB : 759,238	ShapeIT2	Impute2	1000 Genomes Project Phase 3 Version 5	MAF<0.005; info score<0.3	PLINK 1.90 beta	PC1-PC2	Info Score
AFTER EU	sample call rate <98%, extreme heterozygosity, sex mismatches, non-European ancestry, cryptic relatedness, duplicates	Illumina HumanCore Exome v1.0/v1.1	Illumina Genome Studio	Call Rate <=95%, HWE Filter 10e-06, INDELS removed, non 1KG variants removed, 40% MAF difference with 1000G, Duplicate SNPs	318,207	ShapeIT2	Minimac3	1000 Genomes Project Phase 3 Version 5 (updated on Oct 20, 2015)	none	EPACTS	PC1-PC5	r²
Amish	age <18, severe chronic disease, call rate <95%, pHWE<10E-6	Affymetrix 500K and 6.0	BRLMM	Sample call rate <95%, pHWE<5E-6, MAF <0.01	397,704	ShapeIT2	Impute2	1000 Genomes Project Phase 1 Release Version 3 ALL (March 2012)	none	MMAP	NA	Info Score
ARIC EA	Of the 9713 genotyped individuals of European ancestry, we excluded 658 individuals based on discrepancies with previous genotypes, disagreement between reported and genotypic sex, one randomly selected member of a pair of first-degree relatives, or outlier based on measures of average DST or >8 SD away on any of the first 10 principal components.	Affymetrix 6.0	Birdseed	call rate <95%, MAF<0.5%, pHWE<10e-5	682,749	ShapeIT2	Impute2	1000 Genomes Project Phase 1 Release Version 3 ALL (March 2012)	none	SNPTEST v2	PC1-PC10	Info Score
ARIC AA	Of the 3,207 genotyped individuals of Africa ancestry, we excluded 336 individuals based on discrepancies with previous genotypes, disagreement between reported and genotypic sex, one randomly selected member of a pair of first-degree relatives, or outlier based on measures of average DST or >6 SD away on any of the first 10 principal components.	Affymetrix 6.0	Birdseed	call rate <95%, MAF<1%, pHWE<10e-5	773,317	ShapeIT2	Impute2	1000 Genomes Project Phase 1 Release Version 3 ALL (March 2012)	none	SNPTEST v2	PC1-PC10	Info Score
ASPS	Ethnic outliers; duplicates; gender mismatch; cryptic relatedness; sample call rate < 98%; excess heterozygosity	Illumina Human610-Quad BeadChip	Illumina	call rate < 98 %; MAF < 1% ; pHWE < 5×10-6	566,930	ShapeIT2	Impute2	1000 Genomes Project Phase 1 Release Version 3 ALL (March 2012)	none	EPACTS (v3.2.6)	PC1-PC4	Info Score
ASPS-Fam	Ethnic outliers; duplicates; gender mismatch; cryptic relatedness; sample call rate < 98%; excess heterozygosity	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed v2	call rate < 98 %;MAF < 5%;pHWE < 1×10-6	501,288	ShapeIT2	Impute2	1000 Genomes Project Phase 1 Release Version 3 ALL (March 2012)	none	EPACTS (v3.2.6)	PC1-PC4	Info Score
BioMe	none	Illumina HumanOmniExpressExome-8 v1.0	BeadStudio	Removed samples: 1. Sample call rate: < 98% 2. Heterozygosity: coefficient < -0.1 or > 0.3 for common variants (MAF>1%) 3. inbreeding coefficient < 0.4 or > 0.9 for rare variants (MAF<1%) 4. MAF = 0 5. HWE < 1x10-5	AA/HIS: 828,109 EA: 688,734	AA/HIS: ShapeIT2 EA: minimac	AA/HIS: IMPUTE2 EA: Michigan Imputation Server	AA/HIS: 1000 Genomes Project Phase 1 Release Version 3 EA: Haplotype Reference Consortium 1.1	none	EPACTS-3.2.6-patched	PC1-PC8	AA/HIS: Info Score EA: r²
CHS AA	Beyond laboratory genotyping failures, participants were excluded if they had a call rate<=95% or if their genotype was discordant with known sex or prior genotyping (to identify possible sample swaps).	Illumina HumanOmni1-Quad_v1 BeadChip	Illumina GenomeStudio	call rate < 97%, HWE P < 10-5, > 1 duplicate error or Mendelian inconsistency (for reference CEPH trios), heterozygote frequency = 0	940,567	no pre-phasing	Impute2	1000 Genomes Project Phase 3	Variants with insufficient effective minor alleles are filtered prior to analysis. This threshold was set at 5 effective alleles. Where effective alleles is defined as MAF*sampleN*2*impQuality.	custom R software	PC1-PC5	r²
CHS EA	European ancestry participants were excluded from the GWAS study sample due to the presence at study baseline of coronary heart disease, congestive heart failure, peripheral vascular disease, valvular heart disease, stroke or transient ischemic attack or lack of available DNA. Beyond laboratory genotyping failures, participants were excluded if they had a call rate<=95% or if their genotype was discordant with known sex or prior genotyping (to identify possible sample swaps).	Illumina 370CNV BeadChip	Illumina BeadStudio	call rate < 97%, HWE P < 10-5, > 2 duplicate errors or Mendelian inconsistencies (for reference CEPH trios), heterozygote frequency = 0, SNP not found in HapMap.	359,592	MaCH	Minimac1	1000 Genomes Project Phase 3	Variants with insufficient effective minor alleles are filtered prior to analysis. This threshold was set at 10 effective alleles. Where effective alleles is defined as MAF*sampleN*2*impQuality.	custom R software	PC1-PC5	r2
Cilento	Gender mismatch	Illumina 370K (n=859) Illumina OmniExpress(n=758)	Illumina BeadStudio	SNPs in common between the two arrays, call rate<95%, MAF<1%.	~190,000	Eagle	Sanger Imputation Service	Haplotype Reference Consortium	none	EPACTS (fixed version febbrary 2017)	NA	Info Score
DECODE	Call rate < 97%	The chip-typed samples were assayed with the Illumina HumanHap 300, HumanCNV 370, HumanHap 610, HumanHap 1M, HumanHap 660, Omni-1, Omni 2.5 or Omni Express bead chips at deCODE genetics	Graphtyper	Yield < 95%, MAF>0.01, HW < 0.001		Inhouse software	Inhouse software, similar to IMPUTE	Icelandic reference panel - variants matched with Haplotype Reference Consortium or 1000 Genomes Project Phase 3	None	Inhouse software	for quantitative traits: BOLT LMM or variance covariance matrix prop. to the kinship matrix / for binary: adj. for county of birth	Info Score
DIACORE	all patients included	Axiom UK Biobank Array	Axiom GT1 in Genotyping Console 4.0	1) Missing phenotype 2) Ancestry not European 3) Relatedness 2nd degree or closer 4) Genetic gender discordant with phenotypic gender 5) Gonosomal aberation 6) Excess of Heterocygosity 7) Low callrate	799,756	ShapeIT2	Minimac1	1000 Genomes Project Phase 3 Version 5	none	epacts 3.2.6	PC1-PC10	r²

Study	Sample QC	Array	Calling	SNP QC	N SNPs	Phasing	Imputation	Reference panel	Extra filter	Software	Covariates	Metric
ESTHER	Quality control was performed according to Nat. Protoc. 2010 Sept.; 5(9): 1564-1573, Anderson et al.: Gender mismatch, sample call rate < 97%, removal of duplicated or related samples, removal of ethnic outliers (Germans only remained), MAF 0.01, GENO 0.05, HWE 0.00001	Illumina Infinium OncoArray-500K BeadChip	GenomeStudio	MAF < 0.01	368,205	ShapeIT	Impute2	1000 Genomes Project Phase 3 Version 5	none	SNPTEST v2.5.2	not required	Info Score
FHS	call rate >97%, sample failures, genotyped sex different from recorded sex, extreme heterozygosity or high Mendelian error rate	Affymetrix GeneChip Human Mapping 500K Array Set® and 50K Human Gene Focused Panel®	Affymetrix BRLMM	call rate ≥97%, pHWE≥1E-6, Mishap p≥1e-9, ≤100 Mendel errors, MAF≥1%	412,053	ShapeIT	MACH	1000 Genomes Project Phase 1 Release Version 3 (March 2012)	none	GWAF	PCs asscoated with trait with p<0.05	r²
FINCAVAS	call rates < 95%, pHWE < 1E-6, sex mismatch, MDS outliers, excess heterozygosity	Illumina HumanCore Exome and Metabochip	Genome Studio	call rate<95%, pHWE<1e-6, monomorphic removed	HCE: 306,474. MC: 155,499	Eagle2	Minimac3	Haplotype Reference Consortium 1.1	None	EPACTS	PC1-PC5	r²
GCKD	Call rate < 97%, failed sex check, outside 2 SD of mean heterozygosity, cryptic relatedness and genetic ancestry outlier	Illumina Omni2.5Exome BeadChip	Illumina GenomeStudio	Exclude SNPs with call rate < 96%, or HWE p < 1E-5, or MAF < 1%	2,337,794	Eagle	Minimac3	Haplotype Reference Consortium 1.1	none	EPACTS	no associated PCs	r²
Geisinger Research (MyCode)	none	Illumina Human Omni express Exome	Illumina's Genotype studio	Removed samples and markers having: 1. IMPUTE2 info score < 0.7 2. Marker call rate < 99% 3. Sample call rate < 90% 4. MAF < 0.01 5. HWE < 1e-07 6. Removed SNPs having insertions and deletions	589,485	SHAPEIT2	Impute2	1000 Genomes Project Phase 1 Release Version 3 ALL (March 2012)	Removed SNPs with info score<0.7	PLATO v0.0.1	not required	Info Score
HANDLS	Ethnic outliers, cryptic relateds, and sex mismatches, call rate < 95%	Illumina 1M genotyping array	Illumina GenomeStudio	MAF < 0.01, HWE pvalue < 1.0E-07, call rate < 95%	907,763	MACH 1.0	Michigan Imputation Server	1000 Genomes Project Phase 3 Version 5	None	EPACTS (v3.2.6)	PCs	r²
HYPERGENES	Ethnic outliers, sex mismatches, related, call rate<95%; Extremes in heterozygosity	Illumina 1M Duo genotyping array	Illumina GenomeStudio	MAF<0.01; Call rate <99%; HWE < 0.00000004	909,532	ShapeIT	Minimac1	1000 Genomes Project Phase 1 Release Version 3 (March 2012)	none	EPACTS (v3.2.6)	PCs	r²
Jackson Heart Study (JHS)	sex mismatches, sample duplications or swaps, sample call rate <95%	Affymetrix 6.0	Birdseed	call rate <95%	868,969	MACH 1.0	Minimac1	1000 Genomes Project Phase 1 Release Version 3 (March 2012), ALL	none	EPACTS (v3.2.6)	PC1-PC10 and kinship matrix for continuous traits	r²
JMICC	sample call rate < 98 %, sex mismatches, related samples (IBD 0.1875), samples not mapping to JPT (1000 genomes)	Illumina HumanOmniExpressExome	GenomeStudio	Call rate < 98%, pHWE <10e-6, MAF < 1 %, exclude SNPs do not match or not present in 1000 Genomes phase 3 reference panel, remove SNPs with allele frequency difference >20% between scaffold and EAS in 1000GP3, remove duplicates	570,162	ShapeIT2	Minimac3	1000 Genomes Project Phase 3	none	EPACTS	PC1-PC5	r²
KORA_F3	check for European ancestry, check for population outlier	Illumina Omni 2.5/Illumina Omni Express	Genome Studio	call rate >97%, missmatch of phenotypic and genetic gender, 5SD from mean heterozygosity rate, comparison with other genotyping of the same individuals (Metabochip, Exome, Omni)	587,981	ShapeIT	Michigan Imputation Server	1000 Genomes Project Phase 3 Version 5	none	EPACTS (v3.2.6)	PC1-PC10	r²
KORA_F4	check for European ancestry, check for population outlier	Affymetrix Axiom	Affymetrix Software	call rate >97%, missmatch of phenotypic and genetic gender, 5SD from mean heterozygosity rate, comparison with other genotyping of the same individuals (Metabochip, Exome, Omni)	508,532	ShapeIT	Michigan Imputation Server	1000 Genomes Project Phase 3 Version 5	none	EPACTS (v3.2.6)	PC1-PC10	r²
Lifelines	call rate <95%; sex mismatch; heterozygosity > 4SD from mean; non-CaucasiansIBS	Illumina Cyto SNP12 v2	GenomeStudio	samples with call rate < 0.8, excess heterozygosity, non-Caucasian ethnicity (as determined by PCA), high relatedness (pi-hat > 0.4) or a gender mismatch; SNPs with MAF < 1%, a HWE p-value ≤10-3, or a callrate < 95%	257,581		Minimac1	1000 Genomes Project Phase 1 Release Version 3 (March 2012)	none	PLINK 1.90 beta	PC1-PC10	r²
MDC-CC	1. bad call rate 2. excess homozygosity 3. failed gender check 4. Related individuals/duplicates 5. Popoulation outliers	Illumina HumanOmniExpressExome BeadChip v. 1.0	GenomeStudio v2011.1	monomorphic, bad call rate (<95%), fail HWE (p<10^-6)	~800,000	ShapeIT2	Impute2	1000 Genomes Project Phase 1 Release Version 3 ALL (March 2012)	none	SNPTEST	PC1-PC10	Info Score
MESA-AFR	Sex discrepancy, duplicates, call rate <95%, pHW <1E-6, heterozygosity, and outliers	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed v2	call rate≥95%, MA>1%	897,979	ShapeIT2	Michigan Imputation Server	1000 Genomes Project Phase 3 Version 5 ALL	none	EPACTS (v3.2.6)	PC1-PC3	r²
MESA-EUR	Sex discrepancy, duplicates, call rate <95%, pHW <1E-6, heterozygosity, and outliers	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed v2	call rate≥95%, MA>1%	897,979	ShapeIT2	Michigan Imputation Server	Haplotype Reference Consortium	none	EPACTS (v3.2.6)	PC1-PC3	r²
MESA-HIS	Sex discrepancy, duplicates, call rate <95%, pHW <1E-6, heterozygosity, and outliers	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed v2	call rate≥95%, MA>1%	897,979	ShapeIT2	Michigan Imputation Server	1000 Genomes Project Phase 3 Version 5 ALL	none	EPACTS (v3.2.6)	PC1-PC3	r²
MESA-EAS	Sex discrepancy, duplicates, call rate <95%, pHW <1E-6, heterozygosity, and outliers	Affymetrix Genome-Wide Human SNP Array 6.0	Birdseed v2	call rate≥95%, MA>1%	897,979	ShapeIT2	Michigan Imputation Server	1000 Genomes Project Phase 3 Version 5 ALL	none	EPACTS (v3.2.6)	PC1-PC3	r²
METSIM	call rate, sex check, duplicate removal, PC outliers	Illumina HumanOmniExpress-12v1	GenomeStudio	call rate<95%, MAF<1%		ShapeIT2	Minimac3	Haplotype Reference Consortium 1.1	none	EPACTS	mixed-model	r²
NESDA	Non-Caucasians, XO and XXY samples, and samples with a call rate <90%, high genome-wide homo- or heterozygosity, excess IBS	Perlegen-Affymetrix 5.0; Affymetrix 6.0 907K	Birdseed	call rate≤95%; MAF<0.01; pHWE<1E-5; ambiguous location or allele with reference; >20% allele frequency difference from reference; ambiguous SNPs with a MAF>35%	378,163	MACH	Minimac3	1000 Genomes Project Phase 1 Release Version 3 ALL (March 2012)	none	EPACTS	PC1-PC3	r²
OGP	sex mismatches, sample duplications or swaps, sample call rate <95%	Affymetrix 500K Gene Chip	BRLMM	call rate < 95 %;MAF < 1%;pHWE < 1×10-6	347,517	BEAGLE	Michigan Imputation Server	Haplotype Reference Consortium	none	EPACTS (v3.2.6)	Genomic Kinship for quantitative traits; First 3 PCs for binary traits	r²
PIVUS	Call rate <95%; sex mismatch; extreme heterozygosity; related individuals; ancestry outliers	Illumina OmniExpress and Metabochip	Genome Studio	call rate <95%, HWE p<10^-6, MAF<1%	738,583	ShapeIT2	Impute4	Haplotype Reference Consortium	info<0.4	SNPTEST	PC1-PC2	Info Score

Study	Sample QC	Array	Calling	SNP QC	N variants	Phasing	Imputation	Reference panel	Post-imp filter	Assoc. software	PCs	Metric
POPGEN	sample call rate < 90 %, sex mismatches, duplicates Samples (IBD 0.185), samples with heterozygosity outside mean +-3SD, samples not mapping to CEU (Hapmap), i.e. outside median +- 3*IQR and samples with batch problems, i.e. outside median +- 3*IQR	Affymetrix Axiom, Affymetrix 6.0, Illumina Immunochip (Beadchip), Illumina Metabochip, Illumina 550k (merged after QC)	Illumina GenomeS tudio or Illumina OpticalI	SNP call rate < 5%, HWE < 1x10^-5, no MAF for QC but MAF pre Imputation	1049248	ShapeIT2	Impute2	1000 Genomes Project Phase 1 Release Version 3 ALL (March 2012)	removed SNPs with info <= 0.3	EPACTS	not required	Info Score
PREVEND	call rate <95%; sex mismatch; non-Caucasians; duplicated samples	Illumina Cyto SNP12 v2	Illumina GenomeS tudio	call rate < 95%; MAF <1%; pHWE< 1E-4	232571	ShapeIT2	Michigan Imputation Server	Haplotype Reference Consortium	none	SNPTEST V2	PC1-PC5 and exclusion of PC outliers	Info Score
RS-I	MAF < 0.05, SNP callrate < 0.95 and/or HWE p-value < 1 x 10-7, excess heterozygosity, gender swaps, genetic ancestry and familial relationships	Illumina 550K	GeneCall	MAF < 0.05, SNP callrate < 0.95 and/or HWE p value < 1 x 10-7	502668	MaCH	Minimac 3	Haplotype Reference Consortium 1.0	none	RVTEST	PC1-PC5	r²
RS-II	MAF < 0.05, SNP callrate < 0.95 and/or HWE p-value < 1 x 10-7, excess heterozygosity, gender swaps, genetic ancestry and familial relationships	Illumina 550K	GeneCall	MAF < 0.05, SNP callrate < 0.95 and/or HWE p value < 1 x 10-8	490409	MaCH	Minimac 4	Haplotype Reference Consortium 1.0	none	RVTEST	PC1-PC5	r²
RS-III	MAF < 0.05, SNP callrate < 0.95 and/or HWE p-value < 1 x 10-7, excess heterozygosity, gender swaps, genetic ancestry and familial relationships	Illumina 610K and 660K	GeneCall	MAF < 0.05, SNP callrate < 0.95 and/or HWE p value < 1 x 10-9	517658	MaCH	Minimac 5	Haplotype Reference Consortium 1.0	none	RVTEST	PC1-PC5	r²
SHIP	duplicate samples (by IBS), reported/genotyped gender mismatch, callrate <= 92%	Affymetrix SNP 6.0	Birdseed2	pHWE <= 0.0001 or CallRate <= 0.95 or monomorphic SNPs, duplicate IDs, inconsistent reference alleles, mapping problem to build 37	823635	Eagle2	Minimac3	Haplotype Reference Consortium 1.1	none	EPACTS-3.2.6-patched	not required	r²
SiMES	monomorphic, call rate <95%, pHW <1E-6, heterozygosity, related individuals/duplicates, discordant ethnicity, and gender discrepancy.	Illumina Human610-Quad Beadchips	Genomest udio GenTrain and GenCall	T2D DIAMANTE protocol: exclude SNPs do not match or not present in 1000 Genomes phase 3 reference panel, remove SNPs with allele frequency difference >20% between scaffold and reference population in 1000Gp3, remove duplicates	549947	ShapeIT	Michigan Imputation Server	1000 Genomes Project Phase 3 Version 5 ALL	none	EPACTS (v3.2.6)	PC1, PC2	r²
SINDI	monomorphic, call rate <95%, pHW <1E-6, heterozygosity, related individuals/duplicates, discordant ethnicity, and gender discrepancy.	Illumina Human610-Quad Beadchips	Genomest udio GenTrain and GenCall	T2D DIAMANTE protocol: exclude SNPs do not match or not present in 1000 Genomes phase 3 reference panel, remove SNPs with allele frequency difference >20% between scaffold and reference population in 1000Gp3, remove duplicates	552278	ShapeIT	Michigan Imputation Server	1000 Genomes Project Phase 3 Version 5 ALL	none	EPACTS (v3.2.6)	PC1-PC3	r²
SOLID-TIMI 52	individuals excluded if call rate <97%, >3rd degree relative determined by kindship coefficient estimates from KING, GWAS gene didn't match annotated gender	Axiom® Biobank Plus Genotyping Array		call rates <95%, monomorphic, Hardy-Weinberg <E-6,	~547000		HAPI-UR	1000 Genomes Project Phase 1 Release Version 3 ALL (March 2012)	none	EPACTS (v3.2.6)	PC1-PC10	Info Score
STABILITY	individuals excluded if call rate <95%, >3rd degree relative determined by kindship coefficient estimates from KING, GWAS gene didn't match annotated gender	Illumina HumanOmn iExpressExo me-8 v1 array		call rates <95%, monomorphic, Hardy-Weinberg <E-7,	881788	ShapeIT2	Minimac3	1000 Genomes Project Phase 1 Release Version 3 ALL (March 2012)	none	EPACTS (v3.2.6)	PC1-PC10	r²
ULSAM	Call rate <95%; sex mismatch; extreme heterozygosity; related individuals; ancestry outliers	Illumina 2.5M and Metabochip	Genome Studio	call rate <95%, HWE p<10^-6, MAF<1%	1621481	ShapeIT2	Impute4	Haplotype Reference Consortium	info<0.4	SNPTEST	PC1-PC2	Info Score
Vanderbilt-660	sex check, duplicate removal, call rate (<98%), HapMap concordance check	Illumina 660W	Genome Studio	call rate <98%, HWE<0.001, MAF <0.001	527715	ShapeIT	Minimac3	Haplotype Reference Consortium 1.1	none	EPACTS	PC1-PC3	r²
Vanderbilt-AA1M	sex check, duplicate removal, call rate (<98%), HapMap concordance check	Illumina 1M	Genome Studio	call rate <98%, HWE<0.001, MAF <0.001	784048	ShapeIT	Minimac3	Haplotype Reference Consortium 1.1	none	EPACTS	PC1-PC3	r²
Vanderbilt-Omni1	sex check, duplicate removal, call rate (<98%), HapMap concordance check	Illumina OMNI-Quad	Genome Studio	call rate <98%, HWE<0.001, MAF <0.001	924162	ShapeIT	Minimac3	Haplotype Reference Consortium 1.1	none	EPACTS	PC1-PC3	r²
Vanderbilt-Omni5	sex check, duplicate removal, call rate (<98%), HapMap concordance check	HumanOmn i5-Quad	Genome Studio	call rate <98%, HWE<0.001, MAF <0.001	3702007	ShapeIT	Minimac3	Haplotype Reference Consortium 1.1	none	EPACTS	PC1-PC3	r²
YFS	call rates < 95%, pHWE < 1E-6, sex mismatch, MDS outliers, excess heterozygosity	Illumina 670k Custom	Illuminus	call rate<95%, pHWE<1e-6, monomorphic removed	542086	Eagle2	Minimac3	Haplotype Reference Consortium 1.1	None	EPACTS	PC1-PC5	r²
AugUR	sex check, duplicate removal, relatedness, call rate (<98%), HapMap concordance check	Infinium® Global Screening Array-24 v1.0	GenomeS tudio	call rate<95%, pHWE<1e-6, monomorphic removed, removed variants not in reference	614130	ShapeIT	Minimac3	1000 Genomes Project Phase 3 Version 5 ALL	None	rvtests	PC1-PC4	r²
HUNT	Only Europeans were included for this analysis. Samples with call rate <99%, departures from HWE, duplicates, gender mismatch, unusual XY composition, mismatch with reference genome, and samples with contamination > 2.5% were removed	Illumina Hum	GenCall fro	call rate <95%, MAF<0.5%, pHWE<10e-5	368139	Eagle2	Minimac3	Haplotype Reference consortium release 1.1 + 2,201 low-coverage whole-genome sequences samples from the HUNT study	r²≥0.3	SAIGE v0.39	PC1-PC4	r²
MGI	Only European individuals were used for analysis, duplicates, gender mismatch, unusual XY composition, related samples, and samples with contamination > 2.5% were removed	the Illumina Infinium CoreExome-24	GenomeS tudio	Sample call rate < 99%, chromosomal call-rate drop > 5%	502255	Eagle	Minimac3	HRC	none	rvtests	PC1-PC4	r²
UKBB	variants showing batch effects, plate effects, departures from HWE, sex effects, array effects, discordance across control replicates. Samples: ancestry outliers, outliers for heterozygosity and missingness. Further QC details can be found here : https://www.biorxiv.org/content/early/2017/07/20/166298	UK BiLEVE Axiom array, UK Biobank Axiom array	Axiom GT1 algorithm as implement ed in the Affymetrix Power Tools software	Failed QC in > 1 batch, call rate < 95%, MAF < 0.0001, further details can be found here : https://www.biorxiv.org/content/early/2017/07/20/166298	670739	ShapeIT3	Impute4	Haplotype Reference Consortium	None	Quicktest	PC1-PC10	Info Score

[1] References for cited software: MACH (PMID: 19715440); ShapeIT (PMID: 22138821); Eagle (PMID: 27270109); Beagle (PMID: 21310274).
[2] References for cited software: ImputeV2 (PMID: 19543373); minimac3 (PMID: 27571263); PBWT (PMID: 24413527); Sanger Imputation server (PMID: 27548312); Michigan Imputation Server (PMID: 27571263).
[3] References for cited software: EPACTS (Kang, H.M. Epacts: Efficient and Parallelizable Association Container Toolbox. University of Michigan: Department of Biostatistics and Center for Statistical Genetics (2012); PMID: 20208533); SNPTest (PMID: 20517342); RegScan (PMID: 24008273); RVTESTS (PMID: 27153000); PLINK 1.90 (PMID: 25722852); GenABEL (PMID: 17384015); ProbABEL (PMID: 20233392); GWAF (PMID: 20040588); GEMMA (PMID: 22706312); mach2qtl (PMID: 21058334).

Supplementary Table S3: Description of participating studies: phenotype distribution

Study	Subgroup	Ancestry (EA/AA/HIS/EAS/SA)	Time to followup median [years]	Male %	Diabetes at baseline %	Age at baseline median	Age at baseline mean (SD)	eGFRcrea at baseline median (Q1, Q3)	n decline overall	DM at baseline	CKD at baseline
ADVANCE	5	EA	4.35	70%	100%	67.2	67.4 (6.6)	72.1 (59.5, 86.3)	752	752	192
	6	EA	4.35	62%	100%	67.2	67.4 (6.6)	74.0 (62.8, 86.1)	2,169	2,169	436
	UKB	EA	4.35	59%	100%	68.4	67.4 (6.6)	69.3 (57.4, 83.2)	1,061	1,061	319
AFTER EU		EA	6.00	57%	100%	42.7	43.7 (11.1)	89.7 (67.0, 103.9)	831	831	140
Amish		EA	7.00	50%	1%	48.0	48.3 (16.3)	100.4 (88.9, 111.6)	798	NA	NA
ARIC	AA	AA	8.38	37%	20%	53.3	53.9 (5.8)	115.0 (102.8, 123.9)	1,903	298	NA
	EA	EA	8.69	47%	9%	54.6	54.8 (5.7)	101.1 (94.2, 107.4)	7,284	545	NA
ASPS		EA	1.00	43%	0%	65.0	65.8 (8)	73.6 (63.7, 88.1)	469	NA	NA
ASPS-Fam		EA	4.00	40%	0%	68.0	64.6 (10.6)	76.6 (65.3, 86.8)	104	NA	NA
BioMe	Omni EA	EA	2.77	35%	5%	62.9	63.8 (8.7)	76.3 (63.8, 89.1)	852	110	134
	Omni AA	AA	5.34	52%	3%	47.0	47.1 (13.7)	96.6 (79.9, 114.8)	1,717	NA	153
	Omni HA	HIS	4.97	37%	6%	48.4	48.7 (14.8)	92.5 (77.0, 106.1)	2,123	123	180
CHS	AA	AA	4.00	39%	24%	72.0	72.9 (5.7)	72.0 (59.5, 87.2)	481	NA	100
	EA	EA	6.00	44%	12%	71.0	72.3 (5.4)	65.2 (55.3, 75.9)	2,080	210	673
Cilento		EA	8.00	44%	10%	53.0	52.6 (19.7)	92.2 (80.2, 107.1)	788	NA	NA
DECODE		EA	14.00	47%	5%	44.0	45.4 (18.9)	94.1 (78.56, 108.9)	117,666	9,471	10,086
DIACORE		EA	2.96	60%	100%	66.7	65.5 (8.8)	82.4 (67.8, 92.9)	2,169	2,169	352
ESTHER		EA	5.00	42%	17%	62.0	61.6 (6.5)	93.0 (76.5, 103.0)	1,090	155	NA
FHS		EA	15.00	47%	6%	54.0	54.0 (14.9)	74.4 (47.1, 102.1)	2,925	195	1,296
FINCAVAS		EA	8.90	61%	13%	57.8	55.1 (13.2)	90.8 (78.4, 100.0)	835	123	NA
GCKD		EA	2.00	60%	35%	63.0	60.1 (12)	46.4 (37.1, 57.4)	3,941	1,341	3,115
Geisinger Research (MyCode)		EA	13.00	42%	13%	50.0	49 (15.2)	95.1 (80.1, 107.6)	36,286	4,659	2,237
HANDLS		AA	5.00	44%	18%	49.0	48.5 (9)	102.6 (87.6, 116.4)	735	135	NA
HYPERGENES	controls	EA	1.50	61%	0%	57.5	59.5 (9.8)	87.7 (76.9, 97.5)	461	NA	NA
Jackson Heart Study (JHS)		AA	6.60	38%	22%	55.5	55.1 (12.8)	96.5 (80.6, 110.0)	2,162	418	NA
JMICC		EAS	5.03	40%	3%	54.3	54.0 (9.4)	102.2 (96.0,108.4)	975	NA	NA
KORA	F3	EA	10.00	47%	2%	47.0	47.3 (13.0)	104.4 (94.0, 113.8)	2,878	NA	NA
	F4	EA	7.00	49%	3%	49.0	49.2 (13.9)	93.9 (81.9, 105.2)	2,916	NA	NA
Lifelines		EA	5.50	42%	3%	47.0	48.1 (11.4)	94.2 (83.1, 104.1)	10,553	322	142
MDC-CC		EA	16.49	41%	4%	56.3	56.4 (5.7)	80.7 (70.9, 90.6)	2,889		
MESA	AFR	AA	4.00	46%	17%	63.0	62.3 (10.1)	82.3 (70.1, 95.1)	1,283	198	122
	EAS	EAS	4.00	49%	13%	62.0	62.7 (10.2)	83.2 (71.3, 93.7)	615	NA	NA
	EUR	EUR	4.00	48%	5%	63.0	62.4 (10.4)	75.4 (65.6, 86.2)	2,199	128	297
	HIS	HIS	4.00	48%	17%	61.0	61.4 (10.3)	84.2 (71.1, 94.3)	1,176	187	NA
METSIM		EA	4.00	100%	13%	57.0	57.7 (7.1)	93.5 (85.3, 100.0)	5,349	596	NA
NESDA		EA	6.00	34%	4%	43.0	41.9 (13.1)	103.7 (93.9, 114.8)	1,758	NA	NA
OGP		EA	6.34	33%	7%	51.7	53.2 (17.7)	73.1 (61.5, 85.0)	407	NA	NA
PIVUS		EA	5.13	50%	11%	70.1	70.2 (0.2)	81.7 (67.4, 90.6)	539	NA	NA
POPGEN		EA	6.00	53%	3%	57.0	54.7 (14.2)	91.0 (80.0, 100.9)	821	NA	NA
PREVEND		EA	4.00	52%	4%	49.0	49.6 (12.5)	84.3 (73.7, 94.4)	2,932	105	149
RS	I	EA	7.22	40%	13%	72.3	73.2 (7.6)	74.5 (64.3, 84.2)	1,338	121	116
	II	EA	9.30	46%	12%	62.0	64.8 (8.0)	81.7 (71.5, 91.1)	1,248	NA	NA
	III	EA	5.34	46%	9%	56.9	57.2 (6.9)	86.5 (76.9, 95.5)	2,289	NA	NA
SHIP	1	EA	3.00	48%	9%	55.0	54.5 (15.3)	90.4 (75.9, 103.8)	2,163	133	NA
SiMES		EAS	3.67	49%	31%	58.8	59.6 (11.0)	79.3 (64.7, 92.4)	1,451	405	191
SINDI		EAS	4.68	51%	40%	56.8	58 (10.0)	93.9 (80.5, 103.0)	1,554	552	NA
SOLID-TIMI 52	EA	EA	2.00	75%	26%	64.0	64.5 (9.3)	78.9 (65.1, 91.0)	5,759	1,473	938
	EAS	EAS	3.00	83%	34%	65.0	64.7 (9.0)	84.3 (70.1, 92.1)	235	NA	NA
	SA	SA	1.00	79%	34%	62.0	61.0 (11.1)	76.9 (62.4, 92.9)	207	NA	NA
STABILITY	EA	EA	3.00	82%	37%	65.0	64.7 (9.1)	73.4 (61.2, 85.6)	7,687	2,821	1,677
	EAS	EAS	3.00	78%	43%	65.0	64 (9.1)	83.1 (68.1, 92.8)	523	222	NA
	SA	SA	3.00	84%	41%	59.0	58.5 (10.4)	78.5 (64.7, 90.6)	469	NA	NA
ULSAM		EA	6.75	100%	10%	71.0	71.0 (0.6)	57.7 (51.7, 63.9)	686	NA	424
Vanderbilt	660	EA	8.81	45%	4%	54.8	54.0 (15.8)	86.5 (71.1, 100.4)	1,429	NA	146
	AA1M	AA	9.76	34%	6%	47.5	47.6 (16.2)	100.3 (79.6, 119.3)	755	NA	NA
	Omni1	EA	8.97	53%	3%	55.4	54.0 (15.8)	88.2 (70.1, 102.8)	1,859	NA	244
	Omni5	EA	4.32	55%	8%	55.9	52.9 (17.2)	88.2 (69.4, 103.6)	508	NA	NA
YFS		EA	6.00	46%	0%	33.0	31.6 (5.0)	109.7 (100.2, 116.4)	1,683	NA	NA
AugUR		EA	2.40	55%	22%	76.7	77.6 (5.0)	68.9 (59.0, 80.3)	677	147	184
HUNT		EA	21.20	45%	5%	44.5	45.1 (13.7)	104.0 (92.7, 114.2)	46,328	2,235	502
MGI		EA	6.00	46%	39%	52.0	50.4 (15.5)	92.8 (77.7,105.6)	20,077	3,254	1,867
UKBB		EA	4.00	50%	4%	58.0	57.1 (7.3)	92.6 (83.1,99.0)	15,442	542	241

AA: African American ancestry; EA: European ancestry; HIS: Hispanics; SA: South Asian ancestry; EAS: East Asian ancestry

CKD=Chronic Kidney Disease: eGFRcrea at baseline < 60 mL/min/1.73m²

Supplementary Table S4: The 12 identified variants for eGFR-decline were associated with other kidney phenotypes, but not with DM-status. For the 12 identified variants, we show association results for eGFR based in cystatin C[S19]("eGFRcys", n up to 460,826), blood urea nitrogen[S12] ("BUN", n up to 416,178), urine albumin-to-creatinine ratio[S20] ("UACR", n up to 564,257), chronic kidney disease[S12] ("CKD", n up to 625,219) and DM[S21] (n up to 898,130) from published GWAS. Coded allele is the faster-decline allele (which is always the eGFR-lowering allele). Genome-wide significant P-values (P<5.00x10^{-8}) are stated in bold.

SNPID	Locus Name	EA/OA	eGFRcys		BUN		UACR		CKD		DM	
			Beta	P	Beta	P	Beta	P	OR	P	OR	P
Variants with genuine association for eGFR-decline												
rs34882080	*UMOD-PDILT*	a/g	-0.011	**3.44x10^{-75}**	0.010	**4.56x10^{-20}**	-0.011	1.14x10^{-05}	1.205	**3.89x10^{-56}**	0.992	0.570
rs77924615	*UMOD-PDILT*	g/a	-0.012	**6.29x10^{-94}**	0.012	**3.71x10^{-42}**	-0.010	7.24x10^{-05}	1.232	**6.66x10^{-86}**	0.989	0.400
rs10254101	*PRKAG2*	t/c	-0.0090	**1.64x10^{-70}**	0.013	**4.52x10^{-43}**	-0.0029	0.191	1.107	**1.21x10^{-25}**	0.986	0.220
rs1028455	*SPATA7*	t/a	-0.0016	9.51x10^{-04}	0.0012	0.105	0.0026	0.213	1.028	7.65x10^{-04}	0.984	0.160
rs1458038	*FGF5*	c/t	-0.0029	**9.45x10^{-09}**	0.0043	**5.99x10^{-09}**	-0.0029	0.182	1.065	**7.36x10^{-14}**	0.978	0.047
rs4930319	*OVOL1*	c/g	-0.0055	**1.26x10^{-29}**	-0.0050	**3.74x10^{-11}**	-0.0038	0.066	1.060	**7.35x10^{-12}**	1.005	0.640
rs434215	*TPPP*	a/g	-0.0044	**3.10x10^{-14}**	0.0034	0.008	0.0034	0.201	1.043	1.28x10^{-03}	0.989	0.380
rs28857283	*C15ORF54*	g/a	-0.0022	5.76x10^{-06}	0.0026	5.20x10^{-04}	-0.0025	0.210	1.050	**1.19x10^{-08}**	0.986	0.200
rs13095391	*ACVR2B*	a/c	0.00020	0.662	0.0006	0.479	-0.0017	0.743	1.022	0.011	0.983	0.180
Variants without genuine association for eGFR-decline												
rs9998485	*SHROOM3*	a/g	-0.0090	**3.94x10^{-83}**	0.0031	7.68x10^{-04}	-0.012	0.023	1.052	**4.48x10^{-08}**	1.000	0.980
rs1047891	*CPS1*	a/c	0.0039	**3.40x10^{-15}**	-0.0068	**1.26x10^{-15}**	-0.019	**4.01x10^{-18}**	1.053	**2.99x10^{-08}**	0.983	0.130
rs2453533	*GATM*	a/c	-1.00x10^{-04}	0.844	1.00x10^{-04}	0.855	-0.013	**4.49x10^{-10}**	1.076	**8.57x10^{-17}**	0.972	0.010

SNPID=Variant identifier on GRCh37, **Locus name**=Nearest Gene, **EA/OA**=Effect allele / other allele, **Beta** and **P**=genetic effect coefficient of association and association P-value, **OR**=odds ratio, **P**=association P-value.

Supplementary Table S5: **The 12 identified variants for eGFR-decline do not show heterogeneity between ancestries and FHS is not an influential study**. We conducted MR-regression to test for heterogeneity between ancestries[S13] and the meta-analyses restricted to European or African American individuals (n=325,840 and 9,038, respectively; sample sizes for other ancestries were small). We also conducted a sensitivity meta-analysis excluding the FHS study (due to an initial uncertainty in the median eGFR-decline, n=2,925) and explored direction-consistency of genetic effects in FHS alone. Shown are the P-values for between-ancestry heterogeneity (P-anc-het) and beta-estimates in mL/min/1.73m² as well as P-values for the sensitivity analyses; significant P-values ($P_{decline} \leq 0.05/12 = 4.17 \times 10^{-3}$) are stated in bold. Among the 12 variants, there was no evidence for between-ancestry heterogeneity (P-anc-het≥0.05). Association estimates excluding FHS were similar to the original analysis estimates (**Table 1**) and FHS-specific estimates were mostly directionally consistent.

SNPID	Locus Name	EA/OA	P-anc-het	European Beta	European P	African American Beta	African American P	All no FHS Beta	All no FHS P	FHS Beta	FHS P
Variants identified with genuine association for eGFR-decline											
rs34882080	*UMOD-PDILT*	a/g	0.06	0.066	**2.36×10^{-31}**	-0.083	0.174	0.065	**9.70×10^{-30}**	0.091	0.112
rs77924615	*UMOD-PDILT*	g/a	0.85	0.074	**5.50×10^{-37}**	0.016	0.836	0.073	**3.77×10^{-37}**	0.16	0.0423
rs10254101	*PRKAG2*	t/c	0.16	0.020	**7.03×10^{-05}**	0.066	0.223	0.020	**4.35×10^{-05}**	0.019	0.710
rs1028455	*SPATA7*	t/a	0.90	0.020	**1.63×10^{-05}**	0.023	0.517	0.020	**1.12×10^{-05}**	0.076	0.124
rs1458038	*FGF5*	c/t	0.23	0.019	**3.79×10^{-05}**	-0.074	0.257	0.020	**3.03×10^{-05}**	-0.030	0.565
rs4930319	*OVOL1*	c/g	0.70	0.014	**2.19×10^{-03}**	0.043	0.426	0.015	**1.37×10^{-03}**	0.045	0.347
rs434215	*TPPP*[§]	a/g	0.33	0.021	**3.80×10^{-04}**	-0.044	0.532	0.021	**5.43×10^{-04}**	0.12	0.119
rs28857283	*C15ORF54*	g/a	0.22	0.021	**3.44×10^{-06}**	0.075	0.0730	0.022	**1.32×10^{-06}**	0.015	0.745
rs13095391	*ACVR2B*	a/c	0.29	0.018	**1.67×10^{-04}**	0.062	0.207	0.017	**1.77×10^{-04}**	NA	NA
Variants without genuine association for eGFR-decline											
rs9998485	*SHROOM3*	a/g	0.65	0.0048	0.242	0.049	0.222	0.0070	0.156	NA	NA
rs1047891	*CPS1*	a/c	0.35	0.0053	0.287	-0.0040	0.930	0.0040	0.482	0.037	0.456
rs2453533	*GATM*	a/c	0.69	0.0045	0.638	0.022	0.651	0.0010	0.785	0.043	0.360

SNPID=Variant identifier on GRCh37, **Locus name**=Nearest Gene, **EA/OA**=Effect allele / other allele, **P-anc-het**=P-value of the test for between ancestry heterogeneity, **beta** and **P**=genetic effect coefficient of association and association P-value. [§] Since the *TPPP* locus lead variant had imputation quality <0.6 in 45% of the studies (median 0.64), we analyzed this locus omitting the imputation quality filter in all studies.

Supplementary Table S6: No influence by DM-adjustment versus no DM-adjustment or by model-based versus formula-based adjustment for baseline eGFR (eGFR-BL) on the 12 variants' association with eGFR-decline. We conducted a validation meta-analysis for the 12 identified variants for eGFR-decline (total n=103,970) to compare models with different covariate adjustment. Shown are beta-estimates and P-values for eGFR-decline DM-adjusted versus DM-unadjusted, and adjusted for eGFR-baseline by model as well as by formula (**Supplementary Note S1**); all models are age- and sex-adjusted. We found no impact by DM-adjustment, but by adjustment for eGFR-BL (when compared to "not adjusted for DM", which is unadjusted for eGFR-BL). For adjustment for eGFR-BL, we found the same association statistics when model-computed versus formula-derived.

SNPID	EA/OA	EAF	Adjusted for DM		Not adjusted for DM		Adjusted for eGFR-BL by model		Adjusted for eGFR-BL by formula	
			beta	P	beta	P	beta	P	beta	P
Variants identified with genuine association for eGFR-decline										
rs34882080	a/g	0.83	0.058	4.86×10^{-15}	0.058	4.60×10^{-15}	0.077	2.40×10^{-27}	0.078	1.06×10^{-28}
rs77924615	g/a	0.19	0.066	1.68×10^{-19}	0.066	1.34×10^{-19}	0.088	7.83×10^{-37}	0.087	4.73×10^{-37}
rs10254101	t/c	0.28	0.016	0.0130	0.016	0.0125	0.032	1.17×10^{-07}	0.031	1.15×10^{-07}
rs1028455	t/a	0.34	0.020	8.23×10^{-04}	0.020	7.76×10^{-04}	0.021	1.87×10^{-04}	0.021	1.99×10^{-04}
rs1458038	c/t	0.35	0.019	1.67×10^{-03}	0.019	1.68×10^{-03}	0.025	1.04×10^{-05}	0.024	1.31×10^{-05}
rs4930319	c/g	0.34	0.013	0.0241	0.013	0.0279	0.025	1.09×10^{-05}	0.025	5.81×10^{-06}
rs434215	a/g	0.33	0.015	0.0395	0.015	0.0414	0.027	9.20×10^{-05}	0.027	8.72×10^{-05}
rs28857283	g/a	0.37	0.019	1.08×10^{-03}	0.019	1.08×10^{-03}	0.026	2.29×10^{-06}	0.026	2.73×10^{-06}
rs13095391	a/c	0.59	0.022	1.56×10^{-04}	0.022	1.26×10^{-04}	0.027	5.70×10^{-07}	0.027	6.41×10^{-07}
Variants without genuine association for eGFR-decline										
rs9998485	a/g	0.53	0.016	6.67×10^{-03}	0.015	7.05×10^{-03}	0.030	4.55×10^{-08}	0.030	1.78×10^{-08}
rs1047891	a/c	0.31	0.010	0.1232	0.010	0.126	0.031	1.66×10^{-07}	0.030	2.53×10^{-07}
rs2453533	a/c	0.40	-0.0027	0.6378	-0.0028	0.631	0.025	4.46×10^{-06}	0.025	2.78×10^{-06}

SNPID=Variant identifier on GRCh37, **EA/OA**=Effect allele / other allele, **EAF**=Effect Allele Frequency, **beta** and **P**=genetic effect coefficient of association and association P-value.

Supplementary Table S7: Associations of *APOL1* risk variants with eGFR-decline in African American and European ancestry. While our data was derived primarily from persons of European ancestry, we explored variants in the *APOL1* gene due to previous reports for chronic kidney disease progression in 8,500 African American individuals[S22]. We conducted GWAS with the additive model for eGFR-decline unadjusted for eGFR-baseline restricted to African Americans (n up to 9,038) or to European ancestry (n up to 325,840). Shown are beta-estimates (in mL/min/1.73m^2), standard errors (SE) and P-values. From 6 previously reported *APOL1* risk variants (the 7th, indel rs71785313, not analyzable here), none was associated with eGFR-decline in African American ancestry (P≥0.05). Interestingly, we detected two yet unreported SNPs near/in *APOL1* suggestively associated with eGFR-decline with P=2.8x10^{-05} and 3.10x10^{-05} in African Americans (effect allele frequency=0.01; monomorphic in European), uncorrelated with the previously reported variants (r^2<0.01).

SNPID	EA/OA	EAF	African American			European		
			Beta	SE	P	Beta	SE	P
rs73885319	a/g	0.77	0.001	0.05	0.98	NA	NA	NA
rs60910145	t/g	0.78	0.002	0.05	0.97	NA	NA	NA
rs4821480	t/g	0.37	-0.011	0.04	0.78	-0.015	0.0142	0.28
rs2032487	t/c	0.37	-0.004	0.04	0.91	-0.010	0.0131	0.45
rs4821481	t/c	0.37	-0.007	0.04	0.85	-0.001	0.0131	0.45
rs3752462	t/c	0.73	0.027	0.04	0.51	-0.006	0.0047	0.24
rs114021047	a/g	0.01	1.034	0.25	2.80x10^{-05}	NA	NA	NA
rs115045136	t/c	0.01	1.027	0.25	3.10x10^{-05}	NA	NA	NA

SNPID=Variant identifier on GRCh37, **EA/OA**=Effect allele / other allele, **EAF**=Effect Allele Frequency, **beta, SE** and **P**=genetic effect coefficient, standard error of association and association P-value.

SUPPLEMENTARY REFERENCES

S1.   Huffman JE. Examining the current standards for genetic discovery and replication in the era of mega-biobanks. *Nat Commun.* 2018;9(1):5054.

S2.   Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet.* 2006;38(2):209-213.

S3.   Coresh J, Turin TC, Matsushita K, et al. Decline in estimated glomerular filtration rate and subsequent risk of end-stage renal disease and mortality. *JAMA - J Am Med Assoc.* 2014;311(24):2518-2531.

S4.   Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med.* 2009;150(9):604-612.

S5.   The 1000 Genomes Project Consortium. An integrated map of genetic variation Supplementary Material. *Nature.* 2012;135:1-113.

S6.   McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016;48(10):1279-1283.

S7.   Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009;5(6):e1000529.

S8.   Fuchsberger C, Abecasis GR, Hinds DA. Minimac2: Faster genotype imputation. *Bioinformatics.* 2015;31(5):782-784.

S9.   Fuchsberger C, Taliun D, Pramstaller PP, Pattaro C. GWAtoolbox: An R package for fast quality control and handling of genome-wide association studies meta-analysis data. *Bioinformatics.* 2012;28(3):444-445.

S10.   Aschard H, Vilhjálmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am J Hum Genet.* 2015;96(2):329-339.

S11.   Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-2191.

S12.   Wuttke M, Li Y, Li M, et al. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nat Genet.* 2019;51(6):957-972.

S13.   Mägi R, Horikoshi M, Sofer T, 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.* 2017;26(18):3639-3650.

S14.   Hastie T, Tibshirani R. Varying-Coefficient Models. *J R Stat Soc Ser B.* 1993;55(4):757-779.

S15.   Wood SN. *Generalized Additive Models: An Introduction with R, Second Edition.*; 2017.

S16.   Gorski M, Jung B, Li Y, et al. Meta-analysis uncovers genome-wide significant variants for rapid kidney function decline. *Kidney Int.* 2021;99(4):926-939.

S17.   Wanner C, Krane V, März W, et al. Randomized controlled trial on the efficacy and safety of atorvastatin in patients with type 2 diabetes on hemodialysis (4D study): Demographic and baseline characteristics. *Kidney Blood Press Res.* 2004;27(4):259-266.

S18.   Winkler TW, Günther F, Höllerer S, et al. A joint view on genetic variants for adiposity differentiates subtypes with distinct metabolic implications. *Nat Commun.* 2018;9(1):1946.

S19. Stanzick KJ, Li Y, Schlosser P, et al. Discovery and prioritization of variants and genes for kidney function in >1.2 million individuals. *Nat Commun.* 2021;12(1):4350.

S20. Teumer A, Li Y, Ghasemi S, et al. Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nat Commun.* 2019;10(1):4130.

S21. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet.* 2018;50(11):1505-1513.

S22. Parsa A, Kao WHL, Xie D, et al. *APOL1* Risk Variants, Race, and Progression of Chronic Kidney Disease. *N Engl J Med.* 2013;369(23):2183-2196.

Extended acknowledgements, study funding information and author contributions

LifeLines group author genetics

LifeLines Cohort Study: Behrooz Z Alizadeh (1), H Marike Boezen (1), Lude Franke (2), Pim van der Harst (3), Gerjan Navis (4), Marianne Rots (5), Harold Snieder (1), Morris Swertz (2), Bruce HR Wolffenbuttel (6), Cisca Wijmenga (2)
  (1) Department of Epidemiology, University of Groningen, University Medical Center Groningen, The Netherlands
  (2) Department of Genetics, University of Groningen, University Medical Center Groningen, The Netherlands
  (3) Department of Cardiology, University of Groningen, University Medical Center Groningen, The Netherlands
  (4) Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, The Netherlands
  (5) Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, The Netherlands
  (6) Department of Endocrinology, University of Groningen, University Medical Center Groningen, The Netherlands

Author contributions

MG, HR, AT, PM, CAB, TB, CP, AK, FK and IMH **wrote the manuscript**. MG, HR, AT, MWu, CAB, AK, and CP **designed the study**. GS, MSc, BOT, TSA, JÄ, SJLB, BBa, BBr,GB, MB, EB, HB, RJC, JChal, CC, MCi, JCo, MHdB, KEc, ME, CF, RTG, VG, CG, DG, SHa, PH, AHi, KeH, BH, HH, KHv, MAI, BJ, MKä, CK, WKo, HKr, BKK, JK, MLa, TL, LL, RJFL, MAL, OM, YM, AMo, GNN, MNai, MLO, MO, AP, SAP, BWJHP, MP, BMP, OTR, RRe, MR, PR, CSa, HSc, RS, BS, MSi, HSn, KStar, KStef, HSto, KStr, HStr, PS, PvdH, UV, KW, LWal, DMW, HW, CW, TWo, MW, QY, MZ, UT, CAB, AK, FK, CP **managed an individual contributing study**. MG, HR, AT, SGr, FG, MWu, TW, YL, GS, JChai, AC, MCo, MF, AHo, KHo, MLi, MSc, BOT, ATi, JW, TSA, PA, MLBig, GB, RJC, JFC, JChal, MLiC, JC, SF, MGh, SGh, DG, PH, EH, SHw, NSJ, BJ, IK, CK, HKr, BK, LAL, LLy, AMa, PPM, NM, AMo, MNak, MN, BN, IMN, TN, JO, AP, SAP, MHP, LMR, MR, KMR, DR, KR, CSc, SSe, KBS, XS, KStan, KStar, SSz, CHLT, LT, JTr, PvdH, PJvdM, NV, MW, QY, LMY, CAB, AK, CP and IMH **performed statistical methods**