An atlas of target genes, variants, tissues and transcriptional pathways for the regulation of serum urate levels in humans

Adrienne Tin*[1,2], Jonathan Marten*[3], Victoria L. Halperin Kuhns *[4], Yong Li*[5], Matthias Wuttke*[5], Holger Kirsten*[6,7], Karsten B. Sieber[8], Chengxiang Qiu[9], Mathias Gorski[10,11], Zhi Yu[1,12], Ayush Giri[13,14], Gardar Sveinbjornsson[15], Man Li[16], Audrey Y. Chu[17], Anselm Hoppmann[5], Bram Prins[18], Teresa Nutile[19], Damia Noce[20], Masato Akiyama[21,22], Massimiliano Cocca[23], Sahar Ghasemi[24,25], Peter J. van der Most[26], Katrin Horn[6,7], Yizhe Xu[16], Christian Fuchsberger[20], Sanaz Sedaghat[27], Saima Afaq[28,29], Najaf Amin[27], Johan Ärnlöv[30,31], Stephan J.L. Bakker[32], Nisha Bansal[33,34], Daniela Baptista[35], Sven Bergmann[36,37,38], Mary L. Biggs[39,40], Ginevra Biino[41], Eric Boerwinkle[42], Erwin P. Bottinger[43,44], Thibaud S. Boutin[3], Marco Brumat[45], Ralph Burkhardt[7,46,47], Eric Campana[45], Archie Campbell[48], Harry Campbell[49], Robert J. Carroll[50], Eulalia Catamo[23], John C. Chambers[28,51,52,53], Marina Ciullo[19,54], Maria Pina Concas[23], Josef Coresh[1], Tanguy Corre[36,37,55], Daniele Cusi[56,57], Sala Cinzia Felicita[58], Martin H. de Borst[32], Alessandro De Grandi[20], Renée de Mutsert[59], Aiko P.J. de Vries[60], Graciela Delgado[61], Ayse Demirkan[27], Olivier Devuyst[62], Katalin Dittrich[63,64], Kai-Uwe Eckardt[65,66], Georg Ehret[35], Karlhans Endlich[25,67], Michele K. Evans[68], Ron T. Gansevoort[32], Paolo Gasparini[23,45], Vilmantas Giedraitis[69], Christian Gieger[70,71,72], Giorgia Girotto[23,45], Martin Gögele[20], Scott D. Gordon[73], Daniel F. Gudbjartsson[15], Vilmundur Gudnason[74,75], Toomas Haller[76], Pavel Hamet[77,78], Tamara B. Harris[79], Caroline Hayward[3], Andrew A. Hicks[20], Edith Hofer[80,81], Hilma Holm[15], Wei Huang[82,83], Nina Hutri-Kähöne[84,85], Shih-Jen Hwang[86,87], M. Arfan Ikram[27], Raychel M. Lewis[4], Erik Ingelsson[88,89,90,91], Johanna Jakobsdottir[74,92], Ingileif Jonsdottir[15], Helgi Jonsson[93,94], Peter K. Joshi[49], Navya Shilpa Josyula[95], Bettina Jung[10], Mika Kähönen[96,97], Yoichiro Kamatani[21,98], Masahiro Kanai[21,99], Shona M. Kerr[3], Wieland Kiess[7,63,64], Marcus E. Kleber[61], Wolfgang Koenig[100,101,102], Jaspal S. S. Kooner[103], Antje Körner[7,63,64], Peter Kovacs[104], Bernhard K. Krämer[61], Florian Kronenberg[105], Michiaki Kubo[106], Brigitte Kühnel[70], Martina La Bianca[23], Leslie A. Lange[107], Benjamin Lehne[28], Terho Lehtimäki[108,109], Lifelines Cohort Study[110], Jun Liu[27], Markus Loeffler[6,7], Ruth J.F. Loos[43,111], Leo-Pekka Lyytikäinen[108,109], Reedik Magi[76], Anubha Mahajan[112,113], Nicholas G. Martin[73], Winfried März [114,115,116], Deborah Mascalzoni[20], Koichi Matsuda[117], Christa Meisinger[118,119], Thomas Meitinger[101,120,121], Andres Metspalu[76], Yuri Milaneschi[122], Million Veteran Program[123], Christopher J. O'Donnell[124], Otis D. Wilson[125], J. Michael Gaziano[126], Pashupati P. Mishra[127], Karen L. Mohlke[128], Nina Mononen[108,127], Grant W. Montgomery[129], Dennis O. Mook-Kanamori[59,130], Martina Müller-Nurasyid[101,131,132,133], Girish N. Nadkarni[43,134], Mike A. Nalls[135,136], Matthias Nauck[25,137], Kjell Nikus[138,139], Boting Ning[140], Ilja M. Nolte[26], Raymond Noordam[141], Jeffrey OConnell[142], Isleifur Olafsson[143], Sandosh Padmanabhan[144], Brenda W.J.H. Penninx[122], Thomas Perls[145], Annette Peters[71,72,101], Mario Pirastu[146], Nicola Pirastu[49], Giorgio Pistis[147], Ozren Polasek[148,149], Belen Ponte[150], David J. Porteous[48,151], Tanja Poulain[7], Michael H. Preuss[43], Ton J. Rabelink[60,152], Laura M. Raffield[128], Olli T. Raitakari[153,154], Rainer Rettig[155], Myriam Rheinberger[10], Kenneth M. Rice[40], Federica Rizzi[156,157], Antonietta Robino[23], Igor Rudan[49], Rico Rueedi[36,37], Daniela Ruggiero[19,54], Kathleen A. Ryan[158], Yasaman Saba[159], Erika Salvi[156,160], Helena Schmidt[161], Reinhold Schmidt[80], Christian M. Shaffer[50], Albert V. Smith[75], Blair H. Smith[162], Cassandra N. Spracklen[128], Konstantin Strauch[131,132], Michael Stumvoll[163], Patrick Sulem[15], Salman M. Tajuddin[68], Andrej Teren[7,164], Joachim Thiery[7,46], Chris H. L. Thio[26], Unnur Thorsteinsdottir[15], Daniela Toniolo[58], Anke Tönjes[163], Johanne Tremblay[77,165], André G. Uitterlinden[166], Simona Vaccargiu[146], Pim van der Harst[167,168,169], Cornelia M. van Duijn[27], Niek Verweij[167], Uwe Völker[25,170], Peter Vollenweider[171], Gerard Waeber[171], Melanie Waldenberger[70,71,101], Lars Wallentin[172,173], Harvey White[174], John B. Whitfield[73], Sarah H. Wild[175], James F. Wilson[3,49], Qiong Yang[140], Weihua Zhang[52,176], Alan B. Zonderman[68], Murielle Bochud[55], James G. Wilson[177], Sarah A. Pendergrass[178], Kevin Ho[179,180], Afshin Parsa[181,182], Peter P. Pramstaller[20], Bruce M. Psaty[183,184], Carsten A. Böger[10,185], Harold Snieder[26], Adam S. Butterworth[186], Yukinori Okada[187,188], Todd L. Edwards[189,190], Kari Stefansson[15], Katalin

Susztak[9], Markus Scholz[6,7], Iris M. Heid[11], Adriana M. Hung**[125,190], Alexander Teumer**[24,25], Cristian Pattaro**[20], Owen M. Woodward**[4], Veronique Vitart**[3], Anna Köttgen**[†1,5]

* Indicates joint contribution

** Indicates joint oversight

[†] Indicates corresponding author

Authors for Correspondence:

Adrienne Tin
Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
Baltimore, Maryland, USA
+1 443-287-4740
atin1@jhu.edu

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

Author affiliations

1 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA
2 Welch Centre for Prevention, Epidemiology and Clinical Research, Baltimore, Maryland, USA
3 Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
4 Department of Physiology, University of Maryland School of Medicine, Baltimore MD, USA
5 Institute of Genetic Epidemiology, Department of Biometry, Epidemiology and Medical Bioinformatics, Faculty of Medicine and Medical Center - University of Freiburg, Freiburg, Germany
6 Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany
7 LIFE Research Centre for Civilization Diseases, University of Leipzig, Leipzig, Germany
8 Target Sciences - Genetics, GlaxoSmithKline, Collegeville (Pennsylvania), USA
9 Smilow Center for Translational Research, Perelman School of Medicine, University of Pennsylvania
10 Department of Nephrology, University Hospital Regensburg, Regensburg, Germany
11 Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany
12 Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore (Maryland), USA

13 Division of Quantitative Sciences, Department of Obstetrics & Gynecology, Vanderbilt Genetics Institute, Vanderbilt Epidemiology Center, Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN, USA
14 Biomedical Laboratory Research and Development, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA
15 deCODE Genetics, Amgen Inc., Reykjavik, Iceland
16 Department of Medicine, Division of Nephrology and Hypertension, University of Utah, Salt Lake City, USA
17 Genetics, Merck & Co., Inc., Kenilworth, New Jersey, USA
18 Strangeways Research Laboratory, University of Cambridge, 2 Worts' Causeway, Cambridge, CB1 8RN, UK
19 Institute of Genetics and Biophysics Adriano Buzzati-Traverso - CNR, Naples, Italy
20 Eurac Research, Institute for Biomedicine (affiliated to the University of Lübeck), Bolzano, Italy
21 Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan
22 Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
23 Institute for Maternal and Child Health - IRCCS Burlo Garofolo, Trieste, Italy
24 Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
25 DZHK (German Center for Cardiovascular Research), Partner Site Greifswald, Greifswald, Germany
26 Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
27 Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands
28 Department of Epidemiology and Biostatistics, Faculty of Medicine, School of Public Health, Imperial College London, London, UK
29 Institute of Public health & social sciences, Khyber Medical University, Pakistan
30 Department of Neurobiology, Care Sciences and Society, Division of Family Medicine and Primary Care, Karolinska Institutet, Stockholm, Sweden
31 School of Health and Social Studies, Dalarna University, Sweden
32 Department of Internal Medicine, Division of Nephrology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
33 Division of Nephrology, University of Washington, Seattle (Washington), USA
34 Kidney Research Institute, University of Washington, Seattle (Washington), USA
35 Cardiology, Geneva University Hospitals, Geneva, Switzerland
36 Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
37 Swiss Institute of Bioinformatics, Lausanne, Switzerland
38 Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa
39 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle (Washington), USA
40 Department of Biostatistics, University of Washington, Seattle (Washington), USA
41 Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy
42 Human Genetics Centre, University of Texas Health Science Centre, Houston (Texas), USA
43 The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York (New York), USA
44 Digital Health Centre, Hasso Plattner Institute and University of Potsdam, Potsdam, Germany
45 University of Trieste, Department of Medicine, Surgery and Health Sciences, Trieste, Italy

46 Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig, Leipzig, Germany
47 Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg, Germany
48 Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
49 Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
50 Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville (Tennessee), USA
51 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore
52 Department of Cardiology, Ealing Hospital, Middlesex UB1 3HW, UK
53 Imperial College Healthcare NHS Trust, Imperial College London, London, UK
54 IRCCS Neuromed, Pozzilli, Italy
55 Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland
56 Institute of Biomedical Technologies, Italy National Research Council, Bresso (Milano), Italy
57 Bio4Dreams - business nursery for life sciences, Bresso (Milano), Italy
58 San Raffaele Research Institute, Milano, Italy
59 Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands
60 Section of Nephrology, Department of Internal Medicine, Leiden University Medical Centre, Leiden, The Netherlands
61 5th Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology, Diabetology), Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
62 Institute of Physiology, University of Zurich, Zurich, Switzerland
63 Department of Women and Child Health, Hospital for Children and Adolescents, University of Leipzig, Leipzig, Germany
64 Centre for Pediatric Research, University of Leipzig, Leipzig, Germany
65 Intensive Care Medicine, Charité, Berlin, Germany
66 Department of Nephrology and Hypertension, Friedrich-Alexander-University Erlangen- Nürnberg (FAU), Germany
67 Department of Anatomy and Cell Biology, University Medicine Greifswald, Greifswald, Germany
68 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Baltimore (Maryland), USA
69 Department of Public Health and Caring Sciences, Molecular Geriatrics, Uppsala University, Uppsala, Sweden
70 Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany
71 Institute of Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany
72 German Center for Diabetes Research (DZD), Neuherberg, Germany
73 QIMR Berghofer Medical Research Institute, Brisbane, Australia
74 Icelandic Heart Association, Kopavogur, Iceland
75 Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland
76 Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
77 Montreal University Hospital Research Centre, CHUM, Montreal, Canada
78 Medpharmgene, Montreal, Canada

79 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, National Institutes of Health, Bethesda (Maryland), USA
80 Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, Austria
81 Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria
82 Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Centre, Shanghai, China
83 Shanghai Industrial Technology Institute, Shanghai, China
84 Department of Pediatrics, Tampere University Hospital, Tampere, Finland
85 Department of Pediatrics, Faculty of Medicine and Life Sciences, University of Tampere, Finland
86 NHLBIs Framingham Heart Study, Framingham (Massachusetts), USA
87 The Centre for Population Studies, NHLBI, Framingham (Massachusetts), USA
88 Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, USA
89 Stanford Cardiovascular Institute, Stanford University, USA
90 Molecular Epidemiology and Science for Life Laboratory, Department of Medical Sciences, Uppsala University, Uppsala, Sweden
91 Stanford Diabetes Research Center, Stanford University, Stanford, USA
92 The Centre of Public Health Sciences, University of Iceland, Reykjavik, Iceland
93 Landspitalinn University Hospital, Iceland
94 University of Iceland, Iceland
95 Geisinger Research, Biomedical and Translational Informatics Institute, Rockville, USA
96 Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland
97 Department of Clinical Physiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
98 Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan
99 Department of Biomedical Informatics, Harvard Medical School, Boston, USA
100 Deutsches Herzzentrum München, Technische Universität München, Munich, Germany
101 DZHK (German Centre for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich, Germany
102 Department of Internal Medicine II - Cardiology, University of Ulm Medical Centre, Ulm, Germany
103 MRC-PHE Centre for Environment and Health, Imperial College London
104 Integrated Research and Treatment Centre Adiposity Diseases, University of Leipzig, Leipzig, Germany
105 Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria
106 RIKEN Centre for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan
107 Division of Biomedical Informatics and Personalized Medicine, School of Medicine, University of Colorado Denver - Anschutz Medical Campus, Aurora (Colorado), USA
108 Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland
109 Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, Tampere University, Tampere, Finland
110 Lifelines Cohort Study
111 The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York (New York), USA

112 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
113 Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, UK
114 Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany
115 Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz, Austria
116 Medical Clinic V, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
117 Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan
118 Independent Research Group Clinical Epidemiology, Helmholtz Zentrum München, German Research Centre for Environmental Health, Neuherberg, Germany
119 Chair of Epidemiology Ludwig- Maximilians- Universität München at UNIKA-T Augsburg, Augsburg, Germany
120 Institute of Human Genetics, Helmholtz Zentrum München , Neuherberg, Germany
121 Institute of Human Genetics, Technische Universität München , Munich, Germany
122 Department of Psychiatry, VU University Medical Centre, Amsterdam, The Netherlands
123 Department of Veterans Affairs. Office of Research and Development, Washington, DC, USA
124 Boston VA Healthcare System, Department of Veterans Affairs. Office of Research and Development, Washington, DC, USA
125 Vanderbilt University Medical Centre, Division of Nephrology & Hypertension, Nashville, TN, USA
126 Massachusetts Veterans Epidemiology Research and Information Center, VA Cooperative Studies Program, VA Boston Healthcare System, Boston (Massachusetts), USA
127 Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland
128 Department of Genetics, University of North Carolina, Chapel Hill (North Carolina), USA
129 University of Queensland, St Lucia, Australia
130 Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, The Netherlands
131 Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental Health, Neuherberg, Germany
132 Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Germany
133 Department of Internal Medicine I (Cardiology), Hospital of the Ludwig-Maximilians-University (LMU) Munich, Munich, Germany
134 Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York (New York), USA
135 Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda (Maryland), USA
136 Data Tecnica International, Glen Echo (Maryland), USA
137 Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany
138 Department of Cardiology, Heart Center, Tampere University Hospital, Tampere, Finland
139 Department of Cardiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine and Life Sciences, Tampere University, Tampere, Finland
140 Department of Biostatistics, Boston University School of Public Health, Boston (Massachusetts), USA
141 Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Centre, Leiden, The Netherlands
142 University of Maryland School of Medicine, Baltimore, USA
143 Department of Clinical Biochemistry, Landspitali University Hospital, Reykjavik, Iceland

144 Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK
145 Department of Medicine, Geriatrics Section, Boston Medical Center, Boston University School of Medicine, Boston (Massachusetts), USA
146 Institute of Genetic and Biomedical Research, National Research Council of Italy, UOS of Sassari, Li Punti (Sassari), Italy
147 Department of Psychiatry, University Hospital of Lausanne, Lausanne, Switzerland
148 Faculty of Medicine, University of Split, Split, Croatia
149 Gen-info Ltd, Zagreb, Croatia
150 Service de Néphrologie, Geneva University Hospitals, Geneva, Switzerland
151 Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK
152 Einthoven Laboratory of Experimental Vascular Research, Leiden University Medical Centre, Leiden, The Netherlands
153 Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland
154 Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland
155 Institute of Physiology, University Medicine Greifswald, Karlsburg, Germany
156 Department of Health Sciences, University of Milan, Milano, Italy
157 ePhood Scientific Unit, ePhood SRL, Milano, Italy
158 Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, USA
159 Molecular Biology and Biochemistry, Gottfried Schatz Research Centre for Cell Signaling, Metabolism and Aging, Medical University of Graz, Graz, Austria
160 Neuroalgology Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy
161 Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University of Graz, Graz, Austria
162 Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of Dundee, Dundee, UK
163 Department of Endocrinology and Nephrology, University of Leipzig, Leipzig, Germany
164 Heart Centre Leipzig, Leipzig, Germany
165 CRCHUM, Montreal, Canada
166 Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands
167 Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
168 Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands
169 Durrer Centre for Cardiovascular Research, The Netherlands Heart Institute, Utrecht, The Netherlands
170 Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany
171 Internal Medicine, Department of Medicine, Lausanne University Hospital, Lausanne, Switzerland
172 Cardiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden
173 Uppsala Clinical Research Centre, Uppsala University. Uppsala, Sweden
174 Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland, New Zealand
175 Centre for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

176 Department of Epidemiology_and_Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK
177 Department of Physiology and Biophysics, University of Mississippi Medical Centre, Jackson (Mississippi), USA
178 Geisinger Research, Biomedical and Translational Informatics Institute, Danville (Pennsylvania), USA
179 Kidney Health Research Institute (KHRI), Geisinger, Danville (Pennsylvania), USA
180 Department of Nephrology, Geisinger, Danville (Pennsylvania), USA
181 Division of Kidney, Urologic and Hematologic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, USA
182 Department of Medicine, University of Maryland School of Medicine, Baltimore, USA
183 Cardiovascular Health Research Unit, Department of Medicine, Department of Epidemiology, Department of Health Service, University of Washington, Seattle (Washington), USA
184 Kaiser Permanente Washington Health Research Institute, Seattle (Washington), USA
185 Department of Nephrology and Rheumatology, Kliniken Südostbayern AG, Regensburg, Germany
186 Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
187 Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences (IMS), Osaka, Japan
188 Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan
189 Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, Nashville, TN, USA
190 Department of Veterans Affairs, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA

Abstract

Elevated serum urate levels correlate with cardio-metabolic traits and can cause gout. Understanding mechanisms that control serum urate levels may help to develop novel gout therapies and provide insights into correlations between serum urate and cardio-metabolic traits. We performed a large-scale trans-ethnic genome-wide study of serum urate among 457,690 individuals and identified 183 loci (147 novel) that improve risk prediction of gout in an independent sample of 334,880 individuals. Urate-associated variants and genes were prioritized through complementary computational approaches including co-localization with gene expression in 47 tissues. Experimental validation showed that *HNF4A*, a transcriptional master-regulator in liver and kidney, increased transcription of the major urate transporter *ABCG2*, and that HNF4A p.Thr139Ile is a functional variant. These results suggest that transcriptional co-regulation of *HNF4A* target genes may contribute to the complex regulation of serum urate levels and the significant genetic correlations we identified between serum urate and numerous cardio-metabolic traits.

Introduction

Serum urate levels reflect a balance between uric acid production and its net excretion via kidney and intestine. Elevated serum urate levels, hyperuricemia, are correlated with components of the metabolic syndrome as well as with cardiovascular and kidney disease. Hyperuricemia can cause kidney stones and gout, the most common form of inflammatory arthritis[1,2]. Gout attacks are a highly painful inflammatory response to the deposition of urate crystals in hyperuricemia, and are a significant cause of morbidity, emergency room visits, and related health care costs[3]. Although gout has become a major public health issue, it is undertreated because of low awareness, inappropriate prescription practices of the most commonly used drug, allopurinol[4] and poor patient adherence[5]. A better understanding of the mechanisms controlling serum urate levels may not only help to develop novel medications to treat and prevent gout, but may also provide insights into regulatory mechanisms shared with urate-associated cardio-metabolic risk factors and diseases.

Serum urate levels have strong heritable component, with a genetic heritability of 30%-60% in diverse populations, after controlling for age and sex[6-11]. Candidate gene and early genome-wide associations studies (GWAS) have identified three genes as major determinants of serum urate levels: *SLC2A9*, *ABCG2*, and *SLC22A12*[7,12-18]. While *SLC2A9* and *ABCG2* harbor common variants of relatively large effect[19], *SLC22A12* contains many rare or low-frequency variants associated with lower serum urate levels[20]. The largest GWAS meta-analyses of serum urate performed to date identified 28 associated genomic loci among European ancestry (EA) individuals[21] and 27 among Japanese individuals[22]. Genes mapping into the associated loci often encode for urate transporters or their regulators in kidney and gut, as well as for genes relevant to glucose and lipid metabolism, central functions of the liver where uric acid is generated.

Previous GWAS efforts of serum urate have not performed statistical fine-mapping coupled to functional annotation and differential gene expression across tissues. Such approaches benefit from expanding publicly available large datasets, and enable the use of novel methods to prioritize target tissues, pathways, as well as potentially causal genes and

variants[23,24]. Here, we perform large-scale trans-ethnic GWAS meta-analyses of serum urate among 457,690 individuals and identify 183 associated loci that improve risk prediction of gout in an independent sample of 334,880 individuals. Through comprehensive data integration, we prioritize target variants, genes, tissues and pathways that contribute to the complex regulation of serum urate levels. Proof-of-principle experimental verification shows that HNF4A, a transcriptional master regulator in the liver and kidney proximal tubule, increases transcription of *ABCG2*, which encodes a major urate transporter, and that the prioritized HNF4A p.Thr139Ile variant is a functional allele. These results validate our prioritization workflow and support the idea that transcriptional co-regulation of HNF4A target genes contributes to the significant genetic correlations we identify between serum urate and numerous cardio-metabolic traits and diseases.

Results

Meta-analyses for discovery and characterization of serum urate-associated loci

Overview

We developed an automated analysis workflow to collect and integrate results from 74 GWAS of serum urate from five ancestry groups participating in the CKDGen Consortium. We carried out trans-ethnic meta-analyses to obtain general insights into the genetic underpinnings of serum urate and gout, and used EA-specific analyses to dissect loci into genes and pathways as well as to identify genetic correlations with other traits and to evaluate gout risk prediction (**Error! Reference source not found.**).

Trans-ethnic meta-analysis identifies 183 loci associated with serum urate

The primary trans-ethnic GWAS meta-analysis included 457,690 individuals (EA, n=288,649; East Asian ancestry [EAS], n=125,725; African Americans [AA], n=33,671; South Asian ancestry [SA], n=9,037; and Hispanics [HIS], n=608). Mean serum urate levels across studies ranged from 4.2 to 7.2 mg/dl (Error! Reference source not found.). Study-specific GWAS of serum urate were

performed based on genotypes imputed using references panels from the 1000 Genomes Project or the Haplotype Reference Consortium (Methods, **Error! Reference source not found.**). Following standardized study-specific quality control and variant filtering, we combined study-specific results using inverse-variance weighted fixed effect meta-analysis. There was no evidence of un-modeled population stratification (LD Score regression intercept=1.01; λ_{GC}=1.04). After additional post-meta-analysis variant filtering, 8,249,849 SNPs were available for downstream analyses (Methods).

We identified 183 loci that contained at least one SNP associated at genome-wide significance ($p \leq 5x10^{-8}$). A locus was defined as +/-500 kb around the index SNP, the SNP with the lowest p-value (**Error! Reference source not found.**, Error! Reference source not found.). Of these loci, 36 contained a SNP reported as an index SNP in previous GWAS of serum urate[13,15,17,18,21,22,25,26]; the remaining 147 ones were considered novel (**Error! Reference source not found.**). Absolute effect sizes ranged from 0.28 mg/dl per effect allele of rs3775947 (known *SLC2A9* locus) to 0.017 mg/dl at rs11940694 (novel *KLB* locus), with small effects on average (mean absolute effect 0.038 mg/dl, SD 0.033). Regional association plots for the 183 loci are shown in **Error! Reference source not found.**.

Using a summary statistics-based approach (Methods), index SNPs at all 183 loci explained an estimated 7.7% of the trait variance, as compared to 5.3% of variance explained when restricting to 26 index variants previously reported from EA populations[21]. In a large participating pedigree-study from the general population, the 183 lead SNPs explained 17% of serum urate genetic heritability (h^2=37%, 95% credible interval: 29%, 45%), which is a substantial increase over the 5% genetic heritability explained by the three major loci *SLC2A9*, *ABCG2* and *SLC22A12* (**Error! Reference source not found.**; Methods).

Characterization of heterogeneity correlated with ancestry

Most trans-ethnic index SNPs showed homogeneous effects, as indicated by the low values of the I^2 statistic (median 2%, interquartile range 0-14%; Error! Reference source not found.**A**). Fourteen of the 183 index SNPs from the primary trans-ethnic meta-analysis showed evidence of ancestry- associated heterogeneity when tested using a meta-regression approach (Methods,

$P_{anc\text{-}het}<2.7x10^{-4}=0.05/183$), all of which had an I^2 value of >25% (**Figure 1**, **Supplementary Table 3A**). Three principal components generated from a matrix of mean pairwise allele frequency differences between studies were sufficient to separate the self-reported ancestry groups (Error! Reference source not found.). The most significant ancestry-associated heterogeneity was observed for the index variant rs3775947 at *SLC2A9* ($P_{anc\text{-}het}=1.5x10^{-127}$), consistent with observed effect size differences in the ancestry-specific meta-analyses (0.34 mg/dl [EA], 0.26 mg/dl [AA], 0.17 mg/dl [EAS], 0.41 mg/dl [HIS], 0.21 mg/dl [SA]) and previous reports of population heterogeneity of genetic effects at this locus[27]. We identified nine significant ($p<5x10^{-8}$) loci using meta-regression that did not overlap with the significant loci from the primary fixed-effects trans-ethnic meta-analysis. Of these, the index SNPs at *SLC2A2* and *KCNQ1* were also genome-wide significant in EAS (**Supplementary Table 3B**). The overwhelming majority of significant loci in this study, however, showed no heterogeneity correlated with ancestry. Results from ancestry-specific meta-analyses of EA, AA, EAS and SA are summarized in **Supplementary Tables 4 to 7**, respectively, as well as in the **Supplementary Information**. The EA-specific meta-analysis identified 123 genome-wide significant loci (**Supplementary Table 4**) and was used for downstream analyses, detailed below.

Sex-stratified meta-analyses of serum urate GWAS

Male sex is a known, strong correlate of serum urate levels and risk factor for gout. We therefore performed secondary, sex-specific meta-analyses of urate to evaluate whether the urate-associated index SNPs showed sex-specific differences. After multiple-testing correction, six of the 183 trans-ethnic index SNPs showed significant effect differences by sex: *SLC2A9*, *ABCG2*, *CAPN1*, *GCKR*, *IDH2*, and *SLC22A12* ($P_{diff}<2.7x10^{-4}=0.05/183$; **Supplementary Table 8**). A formal test for differences in SNP effects on urate levels between men and women across the genome identified significant ($P_{diff}<5x10^{-8}$, Methods) SNPs in *SLC2A9* and *ABCG2* (**Supplementary Figure 5**), both of which have previously been reported[7,14,15,21], and additional SNPs suggestive of sex differences ($P_{diff}<1x10^{-5}$, Error! Reference source not found.).

Epidemiological and Clinical Landscape

Urate-associated SNPs are associated with gout

To assess the relationship of the 183 trans-ethnic index SNPs with the complex disease gout, we investigated their effects in a trans-ethnic meta-analysis of gout from 20 studies with a total of 763,813 participants, including 13,179 with gout (Methods, **Error! Reference source not found.**, **Supplementary Table 1**). Genetic effects were highly correlated (Spearman correlation coefficient 0.87, **Supplementary Figure 6A**), and 55 SNPs were significantly associated with gout ($p<2.7x10^{-4}$, 0.05/183), supporting the causal role of hyperuricemia in gout. In agreement with previous findings[21], the largest odds ratio (OR) for gout was observed at *ABCG2* (rs74904971, OR 2.04, 95% confidence interval [CI] 1.96-2.12, $P=7.7x10^{-299}$). The genetic effect magnitudes were generally higher at lower minor allele frequency (MAF), with the exceptions of a few large-effect SNPs with >10% MAF that mapped into loci encoding urate transporters with known major effects on urate levels: *SLC2A9*, *ABCG2*, and *SLC22A12*[20] (**Supplementary Figure 6B**).

A genetic risk score for urate improves risk prediction for gout

We evaluated whether a weighted urate genetic risk score (GRS) from independent SNPs improved risk prediction of gout when added to demographic information in a large, independent sample of 334,880 individuals from the UK Biobank, including 4,908 with gout (see Methods). Across categories of the urate GRS, gout prevalence increased from 0.1% in the lowest category (3.61-4.17 mg/dl) to 12.9% in the highest category (6.15-6.44 mg/dl; **Figure 2A, Supplementary Table 10**). In comparison to individuals in the most common GRS category (4.74-5.02 mg/dl), the age- and sex-adjusted OR of gout ranged from 0.09 (95% CI 0.02-0.37, $P=7.8x10^{-4}$) in the lowest category to 13.6 (95% CI 7.2-25.7, $P=1.4x10^{-15}$) in the highest category, corresponding to a >100-fold range (**Figure 2B, Supplementary Table 10**). Of note, 3.5% of the population in the highest three categories of the GRS (≥ 5.87 mg/dL) had a greater than 3-fold increased risk for gout compared to the most common GRS category. This effect size is comparable to a modest effect size for a monogenic disease (OR >3)[28], but much more prevalent in the general population.

Risk prediction models were built by regressing gout status on the GRS alone ("genetic model"), on age and sex ("demographic model"), and finally on the GRS adjusting for age and sex

("combined model") in a training sample consisting of 90% of the individuals. These models were then used to predict gout status in the remaining testing set. The genetic model was a moderately accurate predictor of gout status (area under the receiver operating characteristics curve [AUC]=0.68), weaker than the demographic model (AUC=0.79). The combined model led to a statistically significant increase in prediction accuracy (AUC=0.84, DeLong's test Z=-8.43, p-value $<2.2x10^{-16}$; **Figure 2C**). These observations are consistent with the GRS representing a life-long predisposition to higher urate levels. Because the GRS can be calculated from birth, it may have utility in identifying individuals with a higher genetic risk for gout without knowledge of additional information. This could allow compensatory lifestyle choices to be made earlier in life, reducing the risk of developing this highly painful disease.

High genetic correlations of serum urate with multiple cardio-metabolic traits

Serum urate has been positively correlated with many cardio-metabolic risk factors and diseases[29]. We assessed genome-wide genetic correlations between serum urate and 748 complex traits using the EA-specific meta-analysis results and cross-trait LD score regression (Methods). We identified significant ($p<6.6x10^{-5}=0.05/748$]) associations with 214 complex traits or diseases (**Supplementary Table 11**). The highest positive genetic correlation coefficient (r_g) with a non-urate trait was observed with gout ($r_g=0.92$, $p=3.3x10^{-70}$), followed by traits representing components of the metabolic syndrome: triglycerides in small HDL ($r_g=0.50$), HOMA-IR ($r_g=0.49$), and fasting insulin ($r_g=0.45$). Significant positive genetic correlations were also observed for other cardio-metabolic risk factors or diseases, including waist circumference, obesity, and type 2 diabetes (**Figure 3**). The largest negative correlations were observed with HDL-cholesterol related measurements, consistent with observed associations between high HDL levels and lower cardiovascular risk, and with eGFR ($r_g=-0.26$, $p= 1.4x10^{-9}$), consistent with reduced renal urate excretion at lower eGFR. The genome-wide genetic correlations between serum urate and other complex traits and diseases display a remarkable similarity to the observed associations of serum urate levels with cardio-metabolic traits in epidemiological studies[29].

Identification of enriched tissues and pathways

To identify molecular mechanisms and tissues of importance for urate metabolism, and to provide potential clues into the observed genetic correlation with other traits and diseases, we investigated which tissues, cell types and systems may be significantly enriched for the expression of genes mapping into the urate-associated loci (Methods). Based on all SNPs with $P<1x10^{-5}$ from the trans-ethnic meta-analysis, we identified significant enrichment (false discovery rate [FDR] <0.01) for 19 physiological system entries, three tissues, and two cell types (**Supplementary Table 12**). The strongest enrichment was observed for kidney ($P=9.5x10^{-9}$) and urinary tract ($P=9.9x10^{-9}$), both within the urogenital system, consistent with the kidney's prominent role in controlling serum urate concentrations. There were several other significant entries in the endocrine and digestive system including liver, the major site of urate production. Interestingly, a novel significant enrichment was also observed for entries in the musculoskeletal system, specifically for synovial membrane, joint capsule, and joints (**Figure 4A**), the localization of highly painful gout attacks.

We next tested for cell-type groups with evidence for enriched heritability based on cell-type specific functional genomic elements using stratified LD score regression and the EA-specific meta-analysis results to match the ancestry of the LD score estimates (Methods). The strongest heritability enrichment was observed for kidney (11.5-fold), followed by liver (5.39-fold) and adrenal/pancreas (5.37-fold; **Supplementary Table 13**). This approach complemented the gene-expression based approach and also supported kidney and liver as major organs of urate homeostasis. Results were similar when using trans-ethnic meta-analysis summary statistics (data not shown).

Lastly, we tested whether any gene sets were enriched for variants showing association with serum urate in the trans-ethnic meta-analysis at $P<10^{-5}$ (Methods). Significant enrichment (FDR <0.01) was observed for 383 reconstituted gene sets (**Supplementary Table 14**). As many of these contained overlapping groups of genes, we used affinity propagation clustering (Methods) to identify 57 exemplar gene sets (**Supplementary Table 15**), including a prominent group of inter-correlated gene sets related to kidney and liver development, morphology and function (**Figure 4B**). Together, these analyses underscore the prominent role of the kidney and

liver in regulating serum urate concentrations and implicate the kidney as a major target organ for lowering of serum urate.

Prioritization of urate loci using statistical fine-mapping, function annotation, and gene expression

To prioritize targets for translational research, we established a workflow to couple statistical fine-mapping of urate-associated loci to functional annotation and a systematic evaluation of tissue-specific differential gene expression.

Statistical fine-mapping prioritizes candidate SNPs

To identify independent and potentially causal variants, summary statistics-based fine-mapping was performed based on genome-wide significant loci identified in the EA-specific meta-analysis, because the method relies on LD estimates from an ancestry-matched reference panel whose sample size should scale with that of the GWAS (Methods)[30]. Fine-mapping identified 114 independent SNPs (r^2<0.01) in 99 genomic regions. Most regions contained only one independent signal, nine contained two independent SNPs, the *ABCG2* locus contained three and the *SLC2A9* locus four independent SNPs (**Supplementary Table 16**). For each of these 114 independent SNPs, we computed 99% credible sets representing the set of SNPs which collectively account for 99% posterior probability of containing the variant(s) driving the association signal[31]. The 99% credible sets contained a median of 16 SNPs (IQR 6-57), and six of them only a single variant, mapping in or near *INSR*, *RBM8A*, *MPPED2*, *HNF4A*, *CPT1C*, and *SLC2A9* (**Supplementary Table 16**). Among the 28 small credible sets (≤5 SNPs), several mapped in or near genes with an established role in regulating urate levels such as *SLC2A9*, *PDZK1*, *ABCG2*, *SLC22A11*, and *SLC16A9*. These credible sets contain the most supported candidate causal variants based on association signals and greatly reduce the number of candidate variants for experimental follow-up studies.

To further refine the credible set SNPs, we annotated them with respect to their functional consequence and regulatory potential (Methods). Missense SNPs with posterior probabilities >50% for driving the association signals or mapping into small credible sets were

identified in *ABCG2*, *UNC5CL*, *HNF1A*, *HNF4A*, *CPS1*, and *GCKR* (**Supplementary Table 17, Figure 5A**). All missense SNPs except the one in *GCKR* had a CADD score >15 (Methods), thereby directly implicating the affected gene and SNP as potentially causal. In support, functional effects have already been demonstrated experimentally for variants rs2231142 (Gln141Lys) in *ABCG2*, rs742493 (p.Arg432Gly) in *UNC5CL*, and rs1260326 (p.Leu446Pro) in *GCKR* (**Table 1**). Non-exonic variants with posterior probabilities of >90% and mapping into open chromatin in enriched tissues (Methods) were identified in *RBM8A*, *SLC2A9*, *INSR*, *HNF4A, PDZK1*, *NRG4*, *UNC5CL*, and *AAK1* (**Supplementary Figure 7**, **Supplementary Table 17**). When complemented by evidence of differential gene expression, these SNPs may represent causal regulatory variants and their potential effector genes.

Gene prioritization via gene expression co-localization analyses

To systematically assess differential gene expression, we tested for co-localization of the urate association signals with expression quantitative trait loci (eQTL) in *cis* across three kidney tissue resources and 44 GTEx tissues (Methods). High posterior probability for co-localization (H4≥0.8, Methods) supports a trait-associated variant acting through modulation of gene expression in the tissue where co-localization is identified. The eQTLs from the three kidney tissue resources were based on glomerular and tubulo-interstitial portions of micro-dissected kidney biopsies from 187 CKD patients and healthy kidney tissue sections of 96 additional individuals (Methods). We identified high posterior probability for co-localization with 13 genes in kidney tissue (**Figure 6**), the tissue with the strongest enrichment of signals for urate-associated variants. Whereas co-localization of some genes was restricted to kidney (*SLC17A4*, *BICC1*, *UMOD*, *GALNTL5*, *NCOA7*), other genes showed co-localization across tissues (e.g., *ARL6IP5*). The direction of change in gene expression with higher urate levels could vary for the same gene across tissues. For instance, whereas alleles associated with higher serum urate at the *SLC16A9* locus were associated with higher gene expression in kidney, they were associated with lower expression in other tissues such as aorta, pointing towards tissue-specific regulatory mechanisms[32].

Details on each of the 13 candidate genes with high posterior probability of a shared variant underlying the associations with urate and gene expression in kidney are summarized in

Supplementary Table 18. Significant co-localization signals identified across all 47 tissues are illustrated in **Supplementary Figure 8** and revealed additional novel insights such as co-localization or the urate association signal with expression of *NFAT5* in subcutaneous adipose tissue emphasizing its role in adipogenesis[33], or with expression of *PDZK1* in colon and ileum, important sites of urate excretion. Lastly, we investigated whether any EA-specific index SNPs contained in the 99% credible sets or their proxies (r^2>0.8) were reproducibly associated with gene expression in *trans* in whole blood or peripheral blood mononuclear cell, with results presented in the **Supplementary Information** and **Supplementary Table 19.**

***HNF4A* activates *ABCG2* transcription and HNF4A p.Thr139Ile is a functional variant**

We performed proof-of-principle experimental studies to validate the workflow for prioritizing potentially causal genes and variants, as well as to facilitate insights into the observed genetic correlations of urate levels and cardio-metabolic traits. *HNF1A* and *HNFA4* were selected because they were implicated as causal genes, and because they encode for master regulators of transcription in kidney proximal tubule cells and liver, and shared transcriptional regulation across tissues can potentially explain observed genetic correlations[34].

We first tested whether HNF1A and HNF4A affected the transcription of the ABCG2 gene, which encodes for a urate transporter of major importance in humans. *ABCG2* contains both HNF1A and HNF4A binding sites in its promoter region (**Figure 5B**) and represented the locus with the highest risk for gout in our screen. We used a luciferase reporter assay in HEK 293 cells transiently expressing a construct containing the human *ABCG2* promoter (-1285/+362) upstream of the firefly luciferase gene to assess its transactivation by HNF4A and HNF1A proteins (Methods, **Supplementary Figure 9A**). Co-expression of HNF4A significantly increased the *ABCG2* promoter-driven luciferase activity, and the activation was dependent on the transfected *HNF4A* expression vector dose and corresponding levels of HNF4A protein (**Figure 5C, Supplementary Figure 9B**). No increase of luciferase activity occurred with pGL4 vector without the *ABCG2* promoter (**Supplementary Figure 9D and 9E**). Next, we tested the functional relevance of the prioritized missense p.Thr139Ile allele in HNF4A (NM_178849.2, isoform 1, Methods). Its location within the hinge/ DNA binding domain (DBD) (**Figure 5D, Supplementary Figure 9F**, PBD: 4IQR)[35]

supports potentially altered interactions with targeted promoter regions. The isoleucine substitution at HNF4A p.Thr139Ile significantly increased the transactivation of the *ABCG2* promoter and commiserate luciferase activity as compared to the wildtype threonine (**Figure 5E, Supplementary Figures 9C**) without altering HNF4A protein abundance. Thus, HNF4A can activate *ABCG2* transcription, and HNFA4 p.Thr139Ile is a functional variant. Increased activation of the urate excretory protein ABCG2 by the allele encoding the isoleucine residue should result in lower serum urate levels, which is consistent with our observations from the GWAS. Results for HNF1A indicated that the observed association of this locus with serum urate is unlikely to occur via activation of *ABCG2* (**Figure 5C**).

Discussion

This large trans-ethnic GWAS meta-analysis of serum urate levels based on 457,690 individuals represents a four-fold increase in sample size over previous studies[21,22,36] and resulted in the identification of 183 urate-associated loci, 147 of which were novel. A genetic urate risk score led to significant improvements of gout risk prediction in a large independent sample of 334,880 persons, 3.5% of whom had a gout risk comparable to a modest Mendelian disease effect size. Genome-wide genetic correlation analyses suggested a shared genetic component or co-regulation not only with gout, but also a wide range of cardio-metabolic traits and diseases that reflected known observational correlates of serum urate. Tissue- and cell-type specific enrichment analyses supported kidney and liver, the sites of urate excretion and generation, as key target tissues. Comprehensive fine-mapping and co-localization analyses with gene expression across 47 tissues deliver a comprehensive list of target genes and SNPs for follow-up studies, of which we experimentally confirmed *HNF4A* p.Thr139Ile as a functional allele involved in transcriptional regulation of urate homeostasis.

A major challenge of GWAS is to ascertain the causal gene(s) and/or variants driving each association signal in order to gain novel insights into disease-relevant mechanisms and pathways, and to highlight potential avenues to improve disease treatment and prevention. The datasets

675 generated in this study represent an atlas of candidate SNPs, genes, tissues and pathways
676 involved in urate metabolism that will enable a wide range of follow-up studies. Out of the many
677 novel and biologically plausible findings, we highlight three vignettes in which co-localization of
678 the serum urate and tissue-specific gene expression signals provided new insights into urate
679 metabolism: first, co-localization helped to prioritize genes in association peaks that previous
680 GWAS could not resolve: for example, the association signal at chromosome 6p22.2 contains the
681 genes encoding four members of the SLC17 transporter family (*SLC17A1*, *SLC17A2*, SLC17A3, and
682 *SLC17A4*). Systematic testing of co-localization across genes and tissues supported a shared
683 variant underlying the urate association signal and differential gene expression only for *SLC17A4*
684 in kidney, with higher expression associated with higher serum urate. Previous experimental
685 studies have implicated SLC17A4 as a urate exporter in intestine[37], and our data support its yet
686 unappreciated role in urate transport in the human kidney. Second, co-localization with gene
687 expression provided insights into tubular transport processes that are indirectly connected to
688 urate transport: for example, the gene product of the candidate *ARL6IP5* has been shown to
689 modulate activity of the glutamate transporter SLC1A1[38,39], dysfunction of which causes
690 aminoaciduria[40]; and deletion of the candidate *NCOA7* in mice results in distal renal tubular
691 acidosis[41]. Third, it is noteworthy that co-localization of the urate association signal was observed
692 with differential expression of *MUC1*, *BICC1* and *UMOD* in kidney. Rare mutations in all three
693 genes are known to cause cystic kidney diseases[42-44], pointing towards a shared mechanism with
694 respect to their association with urate.

695 Another noteworthy finding from this well-powered study are the significant genetic
696 correlations with many other, especially cardio-metabolic traits, with directions matching
697 expectation from known observational associations[45]. While the almost perfect genetic
698 correlation with gout reflects a causal relationship, other genetic correlations may reflect co-
699 regulation or broader pleiotropic effects. Many of the moderately but significantly correlated
700 traits reflect central (dys-)functions of the liver or their consequences, including carbohydrate
701 metabolism, diabetes and obesity, as well as lipid metabolism. Together, these findings suggest
702 a shared genetic regulation of metabolic processes in the liver, such as urate generation and lipid
703 metabolism, or an indirect effect of hepatic energy metabolism on urate levels via purine

metabolism. Likewise, significant genetic correlations with kidney-related traits such as eGFR may reflect shared regulation of processes in the kidney, the major site of urate excretion. Evidence for co-regulation is supported by the observation that many urate loci that share associations with other metabolic and kidney function traits encode for transcription factors with major roles in these tissues such as *MLXIPL, TCF7L2*, *HNF1A*, *HNF4A*. Another novel candidate discovered in this screen is *KLF10*, encoding for a transcription factor with an important role in the control of hepatic energy metabolism. <here include a sentence about the omnigenic hypothesis that with sufficient power, all genes active in an trait-relevant tissue will be picked up, which could account for the genetic correlations with traits that are also readouts of hepatic or renal metabolism → interpretation of observed pleiotropy as the potential manifestation of co-regulation of processes that occur in the same trait-relevant tissue >

HNF4A is a powerful illustration of the proposed shared genetic regulation of metabolic processes and excretion of resulting waste products in multiple epithelia types. Mutations in *HNF4A* cause maturity onset diabetes of the young (MODY1)[46] reinforcing its critical role in hepatic and metabolic processes, and this study shows that HNF4A also controls the transcription of *ABCG2*, the key urate secretory transporter in both gut and kidney epithelium (PMID 24441388).[47] Intriguingly, the *HNF4A* T139I functional variant described here increases transcription of the ABCG2 transporter and associates with reduced serum urate levels, is located in a region of the HNF4A protein harboring many of MODY1 mutations (ref). Yet, unlike the severe MODY1 missense mutations [R127W, D126Y, and R125W],[35] the T139I does not cause MODY, but instead increases the risk of type 2 diabetes mellitus, possibly through a tissue specific loss of HNF4A's phosphorylation at T139.[46,48] These data point to additional complexity when interpreting shared associations with possible tissue and gene specific role for *HNF4A* mutations in altering metabolic pathways and urate homeostasis.

In the kidney, nuclear HNF4A, indicative of transcriptional activity, is exclusively detected in the proximal tubule cells[49] and has been reported to regulate the expression of SLC2A9 isoform 1[50] and PDZK1[51]. Kidney-specific deletion of HNF4A in mice phenocopies Fanconi renotubular syndrome.[52] Detailed kidney tissues transcriptomic analyses support HNF4A to drive a proximal tubule signature cluster of 221 co-expressed genes including many candidate genes for urate

metabolism and transport[49]. In addition to *HNF4A*, *HNF4G*, and *HNF1A*, ten genes in this cluster also map into urate-associated loci (*A1CF, CUBN, LRP2, PDZK1, SERPINF2, SLC2A9, SLC16A9, SLC17A1, SLC22A12 and SLC47A1*).

Despite many strengths of this study, some limitations warrant mention. The numbers of individuals of ancestries other than European or East Asian were still small, highlighting the value of studying more diverse populations. Focusing on SNPs present in the majority of studies emphasizes those that may be of greatest importance globally over population-specific variants. General limitations of the field that are not specific to our study are that statistical fine-mapping approaches based on summary statistics from meta-analyses cannot clearly prioritize functional variants in regions of very tight LD, as illustrated by the *ABCG2* locus, and are influenced by the presence of results in the individual contributing studies. Moreover, only few regulatory maps from important target tissues such as synovial membrane and kidney are available, but we were able to evaluate differential gene expression in three separate kidney datasets. The generation of additional regulatory and expression datasets across disease states, developmental stages and more cell types in the kidney and other metabolically active organs represents an important research avenue for the future.

In summary, this large-scale genetic association study of serum urate generated an atlas of candidate SNPs, genes, tissues and pathways involved in urate metabolism and its shared regulation with multiple cardio-metabolic traits that will enable a wide range of follow-up studies.

Online Methods

Overview of GWAS methods

We used a distributive model for study-specific GWAS with meta-analyses conducted centrally. An analysis plan was circulated to all participating studies accompanied by custom shell and R scripts for phenotype generation (https://github.com/genepi-freiburg/ckdgen-pheno). Study-specific GWAS were conducted after a centralized review of the phenotype summary statistics. Study-specific GWAS results were checked using GWAtoolbox[53], including p-value inflation, allele frequency distribution, imputation quality, and completeness of genotypes. Custom scripts were used to compare imputed allele frequencies to those of ancestry-matched reference panels and to visualize variant positions. In addition, quality metrics, including genomic control factor[54], were compared across studies for consistency. The participants of all studies provided written informed consent. Each study had its research protocol approved by the corresponding local ethics committee.

Phenotype definition, genotyping and imputation in participating studies

The primary study outcome was serum urate in mg/dL. The laboratory methods for measuring serum urate in each study are reported in **Supplementary Table 1**. Prevalent gout was analyzed as a secondary outcome to examine whether urate-associated SNPs conferred gout risk. Gout cases were ascertained based on self-report, intake of urate-lowering medications, or International Statistical Classification of Diseases and Related Health Problems (ICD) codes for gout (**Supplemental Table 1**).

Each study performed genotyping separately and applied study-specific quality filters prior to phasing and imputation (**Supplementary Table 2**). In each study, haplotypes were estimated using MACH[55], ShapeIT[56], Eagle[57] or Beagle[58]. Imputation of genotypes was conducted using reference panels from the Haplotype Reference Consortium (HRC) version 1.1[59], 1000 Genomes Project (1000G) phase 3 v5 ALL, or the 1000G phase 1 v3 ALL[60] and ImputeV2[61], minimac3[62], PBWT [63], the Sanger[59], or the Michigan Imputation Server[62]. The imputed genetic

781 dosages were annotated using NCBI b37 (hg19). Each study provided an imputation quality for
782 each variant: ImputeV2 info score, the MACH/ minimac RSQ or the SNPTest info score.
783
784 **Study-specific association analysis**
785 Each study performed ancestry-specific association analysis of serum urate by generating age-
786 and sex-adjusted residuals of serum urate and regressing the residuals on SNP dosage levels,
787 adjusting for study-specific covariates such as study centers and genetic principal components,
788 assuming an additive genetic model. Gout was analyzed as a binary outcome adjusting for age,
789 sex, genetic principal components, and study-specific covariates. Software used for these
790 regression analyses were EPACTS (Test *q.emmax* for family based studies and *q.linear* otherwise;
791 <https://genome.sph.umich.edu/wiki/EPACTS>, SNPTest[64], RegScan[65], RVTEST[66], PLINK 1.90[67],
792 Probabel[68], GWAF[6], GEMMA [25], mach2qtl[69] and R. Family-based studies used methods that
793 accounted for relatedness.

794 **Trans-ethnic, ancestry-specific, and sex-stratified meta-analyses**
795 GWAS results from each study were pre-filtered to retain biallelic SNPs with imputation quality
796 score >0.6 and minor allele count (MAC) >10 before inclusion into meta-analysis. Fixed effects
797 inverse-variance weighted meta-analysis was performed using METAL[70] with modifications to
798 output higher precision (six decimal places). Genomic control was applied for each study. The
799 genomic inflation factor λ_{GC}[54] was calculated to assess inflation of the test statistics. For each
800 meta-analysis result (trans-ethnic, ancestry-specific, and sex-specific), we excluded SNPs that
801 were present in <50% of the studies or with a total MAC <400. For ancestry-specific meta-
802 analysis, we additionally excluded SNPs with heterogeneity >95% as indicated by I^2 to remove
803 signals that were driven by a small number of studies within each ancestry. Genome-wide
804 significance was defined as p-value $<5x10^{-8}$. The LD score regression intercept was calculated to
805 assess the evidence for undetected population stratification[71]. Between-study heterogeneity
806 was assessed using the I^2 statistic[72].
807 In the urate trans-ethnic meta-analysis, 8,249,849 of the 40,534,360 autosomal SNPs
808 analyzed by METAL were retained for downstream characterization after post-meta-analysis

809 filtering. Ancestry-specific meta-analyses were conducted for European ancestry (EA), African
810 American, East Asian ancestry, and South Asian ancestry using the same methods and variant
811 filters as the trans-ethnic meta-analysis. In the EA-specific urate meta-analysis, 8,217,339 of the
812 24,830,632 autosomal SNPs analyzed by METAL were retained for downstream analysis; the LD
813 score regression intercept was 1.0.

814 Secondary meta-analyses were performed separately in men and women, using the same
815 analytical approaches. To test for significant difference of association between males and
816 females, we used a two-sample t-test ($m_{Beta} - f_{Beta}$) / (sqrt($m_{SE\hat{}2} + f_{SE\hat{}2}$)), where m_{Beta} and f_{Beta}
817 were beta coefficients in males and females, respectively, and m_{SE} and f_{SE} were the standard
818 errors among males and females, respectively.

819

820 **Initial determination of genome-wide significant loci**

821 For each meta-analysis results, we scanned the results to search for genome-wide significant
822 SNPs (p-value <$5x10^{-8}$) and defined a locus as a +/-500 kb interval containing at least one genome-
823 wide significant SNP and used the SNP with the lowest p-value in the interval as the index SNP.
824 An ancestry-specific locus was defined as a genome-wide significant locus in an ancestry-specific
825 meta-analysis of which the index SNP did not map into within the +/-500 kb intervals of any
826 genome-wide significant loci in the trans-ethnic meta-analysis.

827

828 **Proportion of phenotypic variance explained and estimated heritability**

829 The proportion of phenotypic variance explained by index SNPs was calculated as the sum of the
830 variance explained by each index SNP calculated as: $\beta^2\left(\frac{2p(1-p)}{var}\right)$, where β is the beta coefficient
831 and p is the MAF of the SNP, and var is the phenotypic variance. For this study, we used the
832 variance of the age- and sex-adjusted residuals of serum urate in European-ancestry participants
833 of the ARIC study as the estimate of the phenotypic variance (1.767).

834 Heritability of age- and sex-adjusted urate was estimated using the R package
835 'MCMCglmm'[73] in the Cooperative Health Research In South Tyrol (CHRIS) study,[74] a participating
836 pedigree-based study of EA individuals (186 up-to-5 generation pedigrees, totaling 4373

individual).[75] We estimated: a) overall heritability, b) heritability excluding index SNPs in three major urate loci (*SLC2A9*, *ABCG2*, and *SLC22A12*), and c) heritability excluding index SNPs in all genome-wide significant loci in the present study. These three estimates were obtained for the trans-ethnic and EA meta-analyses results by running 1,000,000 MCMC iterations (*burn in* = 500,000) based on previously described settings.[75] The difference between the overall heritability and the heritability excluding the index SNPs in the present study represents the heritability explained by the significant loci in the present study.

Trans-ethnic meta-regression

Prior to conducting trans-ethnic meta-regression, we applied the same study-specific SNP filters as those in the trans-ethnic meta-analysis using METAL (imputation quality score >0.6 and MAC >10). An additional filter for minor allele frequency (MAF) >0.0025 was also applied to reduce the influence of very rare SNPs that passed the MAC filter in very large studies. Trans-ethnic meta-regression was conducted using the MR-MEGA software package[76], which models ancestry-associated heterogeneity in the allelic effect as a function of principal components (PCs) of a matrix of mean pairwise allele frequency differences between GWAS studies. Due to software requirements, the minimum number of cohorts for each SNP had to be greater than the number of PCs plus two. Consequently, any SNPs that were present in five or fewer cohorts was excluded from this analysis.

The effect and P-value of each SNP on the phenotype was reported after accounting for heterogeneity. Additional P-values were reported per-SNP for heterogeneity correlated with ancestry ($P_{anc\text{-}het}$) and residual heterogeneity ($P_{res\text{-}het}$). Index SNPs from the METAL meta-analysis with $P_{anc\text{-}het}<5x10^{-8}$ in MR-MEGA were considered to have significant ancestry-associated heterogeneity.

Effect of urate-associated index SNPs on gout and risk prediction for gout

To evaluate the association of the trans-ethnic SNPs with the clinical disease gout, we carried out trans-ethnic meta-analyses of gout using METAL with the same study-specific filtering criteria as the urate trans-ethnic meta-analysis. No post-meta-analysis filtering was performed since the trans-ethnic meta-analysis of gout was only used to assess the association between trans-ethnic

urate index SNPs and gout. For the index SNPs in the trans-ethnic meta-analysis of serum urate, we computed the Spearman correlation between their effects on urate and gout.

The association between a genetic urate risk score constructed from the 114 independent serum urate-associated SNPs identified among individuals of EA (see fine-mapping section below) and gout was assessed in a large, independent sample from the UK Biobank (Project 20272)[77]. The sample was filtered to select only those in the White British ancestry subset, removing individuals with a kinship coefficient greater than 0.0313 and cases of sex chromosome aneuploidy or mismatch between genomically-inferred and self-reported sex. Gout cases were identified by self-report at the inclusion visit, and individuals who developed gout afterwards were excluded as controls using gout-specific ICD codes. The final dataset for analysis included 334,880 individuals, of which 4,908 were classified as gout cases.

The genetic risk score (GRS) was constructed as the sum of the additive imputed dosage of the alleles associated with higher urate levels ("risk alleles"), weighted by the genetic effect of the risk allele on serum urate. The sample was divided into ten bins at evenly spaced intervals between the lowest and highest values of GRS. The lowest bin did not contain any gout cases and was therefore combined with its adjacent bin. Gout status was regressed on GRS bin in a logistic model, including age and sex as covariates with the bin containing the largest number of individuals as the reference group.

To investigate the usefulness of the GRS for the prediction of gout, the sample was divided randomly into a training set containing 90% of the sample and a testing set containing the remainder. Logistic regression models were run regressing gout on GRS (genetic model), age and sex (demographic model) and GRS with age and sex (combined model). Each of these models was then used to predict gout status in the testing set and the performance of the model assessed by comparing to true gout status using Area Under Curve (AUC) in a Receiver Operating Characteristic (ROC) curve.

Genetic correlation

To assess the genetic correlation between serum urate and other traits in EA, we conducted cross-trait LD score regression[71] using LD hub[78] with the EA-specific urate meta-analysis results as input. A total of 746 genetic correlation estimates with serum urate were obtained out of 831 GWAS summary results hosted at LD Hub, excluding two previous serum urate GWAS results. For presentation, the 212 significantly correlated traits ($p<6.7x10^{-5}=0.05/746$) were grouped into 9 categories based on the trait names and labels and presented in a Circos plot.

Functional Enrichment

To assess gene-set and tissue enrichment, we performed the Data-Driven Expression Prioritized Integration for Complex Traits analysis (DEPICT) version 1 release 194[79]. DEPICT performs gene set enrichment analysis by testing whether genes in 14,461 reconstituted gene sets were enriched in GWAS-associated SNPs. These reconstituted gene sets were generated based on similarity analysis from co-regulation of gene expression of 77,840 samples, manually curated gene-sets, molecular pathways from protein-protein interaction screening, and gene sets from mouse gene knock-out studies. Tissues and cell type enrichment was conducted in DEPICT by assessing the gene expression levels of the genes in the associated regions in 37,427 samples quantified using the Affymetrix U133 Plus 2.0 Array platform. The tissue and cell types were mapped to 209 MeSH first level terms including physiological systems, tissues and cells.

All variants with urate association p-values $<1x10^{-5}$ in the trans-ethnic meta-analysis results were used as input. Independent index SNPs were identified using Plink 1.9[67] clump command within 500 kb flanking regions and $r^2>0.1$ in the 1000 Genomes phase1 version 3 data excluding the MHC region (chr6:25–35 Mb). False discovery rates (FDRs) were computed using 500 repetitions, and p-values were computed using 5,000 permutations from 500 null GWAS sets adjusting for gene length.

Affinity Propagation Clustering

Affinity propagation clustering (APC)[80] implemented in the R package 'APCluster'[81] was used to further cluster the urate-related network of reconstituted gene sets containing similar combinations of genes with similarity assessed by the probability of the gene's membership in the gene set. DEPICT reports the top ten genes assigned to each gene set with a z-score

representing the probability of that gene's inclusion within the set. This information was converted into a matrix of genes by pathways, where each element contained a z-score. APC was applied to the similarity matrix derived from this data using a tuning parameter of 0.5 as per the package defaults. The algorithm reports a single data point from each cluster as an 'exemplar' which best represents the points within that cluster. A correlation matrix was calculated from Z-score of each gene within the exemplar gene sets.

LD score regression for functional enrichment

Urate heritability enrichment in 10 cell types in EA was assessed using stratified LD score regression[82] with the EA-specific urate meta-analysis results as the input. The 10 cell types were collapsed from 220 cell-type specific annotations for four histone marks: H3K4me1, H3K4me3, H3K9ac, and H3K27ac. Stratified LD score regression estimates the SNP heritability of urate contributed by the SNPs linked to the histone marks in each cell type. The enrichment of a category is defined as the proportion of SNP heritability in that cell type divided by the proportion of SNPs in the same cell type.

Statistical fine-mapping of genome-wide significant loci in European ancestry

To identify potential causal variants in genome-wide significant loci, we perform fine-mapping in EA given that UKBB genotypes were able to serve as the reference panel with sufficiently large sample size[30]. First, we performed quality control on the UKBB genotypes obtained using Application ID 2027, Dataset ID 8974. We excluded individuals who withdrew consent and removed individuals with mismatched reported and genetic sex, variant missingness >5%, and who represented outliers for variant heterozygosity or along the first two principal components from a principal component analysis seeded with the HapMap phase 3 release 2 populations as reference. We retained only one member of each pair of individuals with a, pair-wise identity-by-descent statistic ≥0.1875. Altogether 13,558 individuals with 16,969,363 SNPs were selected as a random subset used as the LD reference panel for fine-mapping.

Second, neighboring loci with correlated index SNPs (r^2 ≥0.2) in genome-wide significant loci from the EA-specific meta-analysis were combined into independent regions. Third, for each independent region, we performed GCTA independent SNP selection with r2 threshold of <0.01

to identify independent signals[83]. If a region had more than one independent SNP, for each independent SNP, we further conducted conditional analysis controlling for all other independent SNPs using GCTA to generate conditional betas and standard errors for calculating posterior probabilities. Finally, in each independent region, posterior probabilities for each SNP being causal were calculated using a Bayesian methods[31] and 99% credible set were formed by including SNPs with 99% posterior probabilities of containing the causal variant(s).

Annotation of the variants in the credible sets

We annotated SNPs in the credible sets for their exonic effect, Combined Annotation Dependent Depletion (CADD) score, and occurrence in DNaseI-hypersensitive sites (DHS) from the Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenomics Consortium projects[84,85]. The exonic effect and CADD score were obtained using SNiPA v3.2 (March 2017)[86]. SNiPA presented the CADD score as PHRED-like transformation of the C score, which was based on CADD release v1.3 downloaded from http://cadd.gs.washington.edu/download. A CADD score of 15 is used to distinguish potential pathogenic variants from background noise in clinical genetics, and represents the median value of all non-synonymous variants in CADD v1.0[87,88].

Co-localization analysis of cis-eQTL and urate-associated loci

Co-localization of gene expression analysis was conducted using EA-specific urate meta-analysis results, *cis*-eQTL results from micro-dissected human glomerular and tubulo-interstitial kidney portions from 187 individuals in the NEPTUNE study[89], as well as 44 tissues in the GTEx Project version 6p release[32]. For each locus, we identified all genes and all tissue gene pairs with reported eQTL within ±100 kb of each GWAS index SNP. The region for each co-localization test was defined as the eQTL *cis* window in the underlying studies[89,90]. We used the default parameters and prior definitions set in the 'coloc.fast' function from the R package 'gtx' (https://github.com/tobyjohnson/gtx), which is an adapted implementation of Giambartolomei's colocalization method[24]. Evidence for co-localization was defined as H4 ≥0.8, which represents the posterior probability that the association with serum urate and gene expression is due to the same underlying variant. In addition, co-localization of serum urate was also performed with

gene expression quantified using RNA sequencing of the healthy tissue portion of 99 kidney cortex samples from the Cancer Genome Atlas (TCGA)[91]. First, all genes that shared eQTL variants with GWAS index SNPs within ±100 kb were extracted. Then the posterior probability of co-localization was assessed including eQTLs within the *cis*-window (±1Mb from the transcription start site) for each gene using the R coloc package[24] with default values for the three prior probabilities.

Trans-eQTL annotation by LD mapping

We performed trans-eQTL annotation by LD mapping using the 1000Genomes, phase 3 European reference for LD with a cut-off of r^2>0.8. The SNPs in this analysis included index SNPs in EA-specific meta-analysis with >1% posterior probability. Due to small effect sizes, only large trans-eQTL studies with sample size >1,000 individuals were considered, namely[92-96] , the latter updated by a larger sample size and combining two studies (LIFE-Heart[97], and LIFE-Adult[98]) with a total sample size of 6,645. To improve stringency of results, we only report inter-chromosomal trans-eQTLs showing gene expression association p-values $<5x10^{-8}$ in at least two of the above mentioned independent sample sets.

Experimental study

Promoter Binding Site Predictions. Using the JASPAR 2018 database[99,100], frequency matrices were downloaded for transcription factor binding sites of both vertebrate and human sequences (HNF1A: MA0046.1 and MA0046.2; HNF4A: MA0114.1 and MA0114.2). These matrices were then used to query the promoter region of *ABCG2* (-1285/+362)[101] by means of the LASAGNA 2.0[102] transcription factor binding site search tool with default parameters and a p value cutoff of 0.01.

Site-Directed Mutagenesis. HNF1A and HNF4A clones were purchased from GeneCopoeia, (EX-A7792-M02 and EX-Z5283-M02 respectively) and were mutagenized using the QuikChange Lightning Site Directed Mutagenesis kit (Agilent Technologies, #210518) per manufacturer's instructions using PAGE purified primers.

(HNF1A-A98V-Forward: CCCTGAGGAGGCGGTCCACCAGAAAGCCG;

HNF1A-A98V-Reverse: CGGCTTTCTGGTGGACCGCCTCCTCAGGG;

HNF4A-T139I-Forward: GACCGGATCAGCATTCGAAGGTCAAGC;

HNF4A-T139I-Reverse: GCTTGACCTTCGAATGCTGATCCGGTC).

Luciferase Assay. HEK293T cells were seeded in white walled 96 well plates coated with Poly-L-lysine at roughly 12,500 cells per well. Cells were transfected 18 hours later with either the ABCG2 promoter (-1285/+362) upstream of a firefly luciferase in the pGL4.14 vector (Promega, #E699A), or the pGL4.14 vector without promoter construct, as well as GFP expressing vector used as an internal control (pEGFP-C1, Clontech)[103] using X-tremeGene™ 9 DNA Transfection Reagent (Roche Diagnostics, #6365787001). Transfection cocktails were prepared per manufacturer's specifications either with or without transcription factor using the following ratio: 0.6 μg promoter construct, 0.2 μg transcription factor, and 0.05 μg GFP. When no transcription factor was used, pcDNA3.1 was substituted, and if more than one transcription factor was used, 0.1 μg of each was used such that the sum of those transcription factors was equal to 0.2 μg DNA. Approximately 48 hours after transfection, cells were rinsed with 1x PBS, then lysed using Passive Lysis Buffer (Promega) for 15 minutes. During this incubation, GFP measurements were taken using a CLARIOstar microplate reader (BMG Labtech). Next, 30 μl of Luciferase Reagent (Promega, E297A&B) were added to each well, and the plate was incubated for an additional 20 minutes at room temperature. Finally, luciferase activity was measured using the CLARIOstar microplate reader taking the average over 6 seconds.

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

Gene	SNP	#SNPs in set	SNP PP	consequence	CADD	DHS	Gout p-value (EA)	Brief summary of literature and gene function
ABCG2	rs2231142	4	0.41	p.Gln141Lys (NP_004818.2)	18.2	ENCODE epithelial	1.21E-290	Encodes a xenobiotic and high-capacity urate membrane transporter expressed in kidney, liver and gut. Causal variants have been reported for gout susceptibility (#138900) and the Junior Jr(a-) blood group phenotype (#614490). The locus was first identified in association with serum urate through GWAS (PMID:18834626) and confirmed in many studies since. The common causal variant Q141K has been experimentally confirmed (PMID:19506252) as a partial loss of function.
UNC5CL	rs742493	4	0.95	p.Arg432Gly (NP_775832.2) (within Death domain)	21.0	ENCODE epithelial	2.73E-01	Encodes for the death-domain-containing Unc-5 Family C-Terminal-Like membrane-bound protein. Suggested as a candidate gene for mucosal diseases, with a role in epithelial inflammation and immunity (PMID:22158417). Experiments using human HEK293 cells showed that UNC5CL can transduce pro-inflammatory programs via activation of NF-□B, with the 432Gly variant less potent to do so than the 432Arg one (PMID:22158417).
HNF1A	rs1800574	2	0.92	p.Ala98Val (NP_000536.5)	23.4		1.83E-02	Encodes a transcription factor with strong expression in liver, guts and kidney. Rare mutations cause autosomal-dominant MODY type III (#600496). Locus found in GWAS of T2DM (PMID:22325160) and blood urea nitrogen (PMID:29403010). Together with HNF4-alpha, it was first recognized as master regulator of hepatocyte and islet transcription. Knockout mice show proximal tubular dysfunction (Fanconi syndrome). HNF1A enhanced promoter activity of PDZK1, URAT1, NPT4 and OAT4 in human renal proximal tubule cell-based assays (PMID:28724612), supporting a role in the coordinated expression of components of the urate "transportosome".
HNF4A	rs1800961	1	1.00	p.Thr139Ile (NP_000448.3)	24.7	ENCODE pancreas	7.43E-03	Encodes another nuclear receptor and transcription factor that controls expression of many genes, including *HNF1A* and other overlapping target genes. Rare mutations cause autosomal-dominant MODY type I (#125850) and autosomal-dominant renal Fanconi syndrome 4 (# 616026). Shown to regulate expression of SLC2A9 and other members of the urate "transportosome" in cell-based assays (PMID 25209865, PMID:30124855). The GWAS locus has been reported for multiple cardio-metabolic traits and T2DM (PMID:21874001).
CPS1	rs1047891	84	0.84	p.Thr1412Asn (NP_001116105.1)	22.1		5.66E-02	Encodes mitochondrial carbamoyl phosphate synthetase I, which catalyzes the first committed step of the urea cycle by synthesizing carbamoyl phosphate from ammonia, bicarbonate, and 2 molecules of ATP. Rare mutations cause autosomal-recessive carbamoylphosphate synthetase I deficiency (#237300). In addition to hyperammonemia, this disease features increased synthesis of glutamine, a precursor of purines. Elevated uric acid excretion has been reported in patients with hyperammonemia (PMID:6771064). GWAS locus for eGFR (PMID:26831199), homocysteine (PMID:23824729), urinary glycine concentrations (PMID: 26352407).
GCKR	rs1260326	2	0.67	p.Leu446Pro (NP_001477.2)	0.1	ENCODE kidney	4.09E-41	Encodes a regulatory protein prominently expressed in the liver that inhibits glucokinase. Identified in previous GWAS of urate (PMID:23263486) and multiple other cardio-metabolic traits. The 446L protein was shown to be less activated than 446Pro by physiological concentrations of fructose-6-phosphate, leading to reduced glucokinase inhibitory ability (PMID:19643913).

Abbreviation: pp, posterior probability; DHS, DNases hypersensitivity site; CADD, Combined Annotation Dependent Depletion phred score; EA, European ancestry.

1025

Acknowledgements

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

Disclaimer

The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Heart, Lung, and Blood Institute, the National Institutes of Health, or the US Department of Health and Human Services.

Data Availability

Genome-wide summary statistics for this study are made publicly available through dbGaP accession number phs000930.v7.p1.

Author contributions

Manuscript writing group: Adrienne Tin, Jonathan Marten, Victoria L. Halperin Kuhns, Yong Li, Matthias Wuttke, Holger Kirsten, Karsten B. Sieber, Chengxiang Qiu, Mathias Gorski, Markus Scholz, Adriana M. Hung, Alexander Teumer, Cristian Pattaro, Owen M. Woodward, Veronique Vitart, Anna Köttgen.

Design of this study: Adrienne Tin, Jonathan Marten, Matthias Wuttke, Mathias Gorski, Christian Fuchsberger, Alexander Teumer, Cristian Pattaro, Owen M. Woodward, Veronique Vitart, Anna Köttgen,

Management of an individual contributing study: Adam S. Butterworth, Adriana M. Hung, Adrienne Tin, Afshin Parsa, Aiko P.J. de Vries, Alan B. Zonderman, Alessandro De Grandi, Andres Metspalu, Andrew A. Hicks, Anke Tönjes, Anna Köttgen, Annette Peters, Antje Körner, Antonietta Robino, Archie Campbell, Belen Ponte, Bernhard K. Krämer, Bettina Jung, Brenda W.J.H. Penninx, Bruce M. Psaty, Caroline Hayward, Carsten A. Böger, Cassandra N. Spracklen, Christian Gieger, Christopher J. O'Donnell, Cornelia M. van Duijn, Cristian Pattaro, Daniela Toniolo, Daniele Cusi, Deborah Mascalzoni, Eric Boerwinkle, Erik Ingelsson, Florian Kronenberg, Gardar Sveinbjornsson, Georg Ehret, Gerard Waeber, Ginevra Biino, Girish N. Nadkarni, Grant W. Montgomery, Harold Snieder, Harvey White, Helena Schmidt, Igor Rudan, J. Michael Gaziano, James G. Wilson, Jaspal S. S. Kooner, Jeffrey OConnell, Joachim Thiery, Johanne Tremblay, John B. Whitfield, John C. Chambers, Josef Coresh, Kai-Uwe Eckardt, Karen L. Mohlke, Kari Stefansson, Kevin Ho, Koichi Matsuda, Konstantin Strauch, Lars Wallentin, M. Arfan Ikram, Marcus E. Kleber, Marina Ciullo, Mario Pirastu, Markus Loeffler, Markus Scholz, Martin H. de Borst, Matthias Wuttke, Michael Stumvoll, Michele K. Evans, Michiaki Kubo, Mika Kähönen, Murielle Bochud, Myriam Rheinberger, Nicholas G. Martin, Olivier Devuyst, Olli T. Raitakari, Ozren Polasek, Paolo Gasparini, Peter P. Pramstaller, Peter Vollenweider, Pim van der Harst, Qiong Yang, Rainer Rettig, Reinhold Schmidt, RenÕ©e de Mutsert, Robert J. Carroll, Ron T. Gansevoort, Ruth J.F. Loos, Sarah A. Pendergrass, Sarah H. Wild, Stephan J.L. Bakker, Tamara B. Harris, Terho Lehtimäki, Thomas Perls, Ton J. Rabelink, Uwe Völker, Vilmantas Giedraitis, Vilmundur Gudnason, Weihua Zhang, Wieland Kiess, Winfried März, Wolfgang Koenig, Yong Li, Yuri Milaneschi

Critical review of manuscript: Adam S. Butterworth, Adriana M. Hung, Adrienne Tin, Afshin Parsa, Aiko P.J. de Vries, Alan B. Zonderman, Albert V. Smith, Alexander Teumer, André G. Uitterlinden, Anke Tönjes, Anna Köttgen, Annette Peters, Anselm Hoppmann, Antje Körner, Antonietta Robino, Anubha Mahajan, Audrey Y. Chu, Ayush Giri, Bernhard K. Krämer, Bettina Jung, Boting Ning, Bram Prins, Brenda W.J.H. Penninx, Brigitte Kühnel, Bruce M. Psaty, Caroline Hayward, Carsten A. Böger, Cassandra N. Spracklen, Chengxiang Qiu, Christa Meisinger, Christian Fuchsberger, Christian Gieger, Christopher J. O'Donnell, Cristian Pattaro, Daniel F. Gudbjartsson, Daniela Ruggiero, Deborah Mascalzoni, Dennis O. Mook-Kanamori, Erik Ingelsson, Erwin P. Bottinger, Eulalia Catamo, Florian Kronenberg, Gardar Sveinbjornsson, Ginevra Biino, Giorgia Girotto, Girish N. Nadkarni, Graciela Delgado, Grant W. Montgomery, Harold Snieder, Harry Campbell, Harvey White, Helgi Jonsson, Hilma Holm, Igor Rudan, Ilja M. Nolte, Ingileif Jonsdottir, Iris M. Heid, James G. Wilson, Johanna Jakobsdottir, Johanne Tremblay, John B. Whitfield, Jonathan Marten, Josef Coresh, Kai-Uwe Eckardt, Karen L. Mohlke, Karlhans Endlich, Karsten B. Sieber, Katalin Susztak, Kenneth M. Rice, Kevin Ho, Kjell Nikus, Konstantin Strauch, Lars Wallentin, Laura M. Raffield, Leo-Pekka Lyytikäinen, Leslie A. Lange, Man Li, Marcus E. Kleber, Marina Ciullo, Markus Loeffler, Markus Scholz, Martin H. de Borst, Martina La Bianca, Martina Müller-Nurasyid, Mary L. Biggs, Mathias Gorski, Matthias Nauck, Matthias Wuttke, Melanie Waldenberger, Michael H. Preuss, Michele K. Evans, Mika Kähönen, Mike A. Nalls, Myriam Rheinberger, Nicholas G. Martin, Niek Verweij, Nina Hutri-Kähöne, Nisha Bansal, Olivier Devuyst, Olli T. Raitakari, Otis D. Wilson, Ozren

Polasek, Patrick Sulem, Pavel Hamet, Peter K. Joshi, Pim van der Harst, Qiong Yang, Rainer Rettig, Ravchel M. Lewis, Raymond Noordam, RenÕ©e de Mutsert, Ruth J.F. Loos, Sahar Ghasemi, Sala Cinzia Felicita, Salman M. Tajuddin, Sanaz Sedaghat, Sarah A. Pendergrass, Sarah H. Wild, Scott D. Gordon, Shih-Jen Hwang, Shona M. Kerr, Stephan J.L. Bakker, Tamara B. Harris, Teresa Nutile, Terho Lehtimäki, Thibaud S. Boutin, Thomas Meitinger, Todd L. Edwards, Ton J. Rabelink, Unnur Thorsteinsdottir, Uwe Völker, Veronique Vitart, Wei Huang, Winfried März, Wolfgang Koenig, Yong Li, Zhi Yu.

Statistical Methods and Analysis: Albert V. Smith, Alexander Teumer, Anna Köttgen, Anselm Hoppmann, Anubha Mahajan, Audrey Y. Chu, Ayse Demirkan, Ayush Giri, Bettina Jung, Boting Ning, Bram Prins, Brigitte Kühnel, Carsten A. Böger, Cassandra N. Spracklen, Chengxiang Qiu, Chris H. L. Thio, Christian Fuchsberger, Cristian Pattaro, Damia Noce, Daniel F. Gudbjartsson, Edith Hofer, Erika Salvi, Federica Rizzi, Gardar Sveinbjornsson, Ginevra Biino, Graciela Delgado, Holger Kirsten, Ilja M. Nolte, Iris M. Heid, James F. Wilson, Johanna Jakobsdottir, Johanne Tremblay, Jonathan Marten, Jun Liu, Karsten B. Sieber, Katalin Susztak, Kathleen A. Ryan, Katrin Horn, Kenneth M. Rice, Laura M. Raffield, Leo-Pekka Lyytikäinen, Leslie A. Lange, Man Li, Marco Brumat, Marcus E. Kleber, Maria Pina Concas, Markus Scholz, Martin Gögele, Mary L. Biggs, Masahiro Kanai, Masato Akiyama, Massimiliano Cocca, Mathias Gorski, Matthias Nauck, Matthias Wuttke, Michael H. Preuss, Mike A. Nalls, Myriam Rheinberger, Navya Shilpa Josyula, Nicola Pirastu, Niek Verweij, Nina Mononen, Pashupati P. Mishra, Pavel Hamet, Peter J. van der Most, Peter K. Joshi, Pim van der Harst, Qiong Yang, Raymond Noordam, Rico Rueedi, Robert J. Carroll, Sahar Ghasemi, Salman M. Tajuddin, Sanaz Sedaghat, Sarah A. Pendergrass, Shih-Jen Hwang, Tanguy Corre, Teresa Nutile, Thibaud S. Boutin, Todd L. Edwards, Toomas Haller, Veronique Vitart, Weihua Zhang, Winfried März, Yasaman Saba, Yizhe Xu, Yoichiro Kamatani, Yong Li, Yukinori Okada

Subject Recruitment: Aiko P.J. de Vries, Alan B. Zonderman, Andrej Teren, Andres Metspalu, Anke Tönjes, Anna Köttgen, Archie Campbell, Belen Ponte, Bettina Jung, Blair H. Smith, Brenda W.J.H. Penninx, Carsten A. Böger, Christa Meisinger, Cristian Pattaro, Daniela Ruggiero, Daniele Cusi, David J. Porteous, Erwin P. Bottinger, Florian Kronenberg, Gerard Waeber, Harry Campbell, Harvey White, Helgi Jonsson, Igor Rudan, Isleifur Olafsson, James G. Wilson, Jaspal S. S. Kooner, Johan Ärnlöv, Johanne Tremblay, John B. Whitfield, John C. Chambers, Katalin Dittrich, Kjell Nikus, Koichi Matsuda, Lars Wallentin, Marina Ciullo, Michele K. Evans, Michiaki Kubo, Mika Kähönen, Myriam Rheinberger, Nicholas G. Martin, Nina Hutri-Kähöne, Olli T. Raitakari, Ozren Polasek, Patrick Sulem, Peter Vollenweider, Reinhold Schmidt, Renée de Mutsert, Ron T. Gansevoort, Saima Afaq, Sandosh Padmanabhan, Sarah A. Pendergrass, Sarah H. Wild, Simona Vaccargiu, Tanja Poulain, Terho Lehtimäki, Ton J. Rabelink, Vilmundur Gudnason, Wei Huang, Winfried März

Bioinformatics: Albert V. Smith, Anna Köttgen, Anselm Hoppmann, Audrey Y. Chu, Ayush Giri, Benjamin Lehne, Bram Prins, Carsten A. Böger, Cassandra N. Spracklen, Chengxiang Qiu, Christian M. Shaffer, Daniela Baptista, Dennis O. Mook-Kanamori, Edith Hofer, Eric Campana, Erika Salvi, Federica Rizzi, Georg Ehret, Giorgio Pistis, Holger Kirsten, Iris M. Heid, James F. Wilson, Johanna Jakobsdottir, Johanne Tremblay, Jonathan Marten, Karen L. Mohlke, Karsten B. Sieber, Katalin Susztak, Katrin Horn, Leo-Pekka Lyytikäinen, Man Li, Marcus E. Kleber, Maria Pina Concas, Markus Scholz, Massimiliano Cocca, Mathias Gorski, Matthias Wuttke, Michael H. Preuss, Navya Shilpa Josyula, Nicola Pirastu, Pashupati P. Mishra, Pavel Hamet, Peter J. van der Most, Raymond Noordam, Reedik Magi, Rico Rueedi, Robert J. Carroll, Sahar Ghasemi, Sanaz Sedaghat, Sarah A. Pendergrass, Scott D. Gordon, Sven Bergmann, Tanguy Corre, Teresa Nutile, Weihua Zhang, Winfried März, Yasaman Saba, Yizhe Xu, Yong Li, Yuri Milaneschi, Zhi Yu.

Interpretation of Results: Adrienne Tin, Alexander Teumer, André G. Uitterlinden, Anna Köttgen, Ayush Giri, Bettina Jung, Carsten A. Böger, Cassandra N. Spracklen, Chengxiang Qiu, Christian Gieger, Christopher J. O'Donnell, Cristian Pattaro, Harvey White, Helgi Jonsson, Holger Kirsten, Iris M. Heid, Johanne Tremblay, Jonathan Marten, Karen L. Mohlke, Karlhans Endlich, Karsten B. Sieber, Katalin Dittrich, Katalin Susztak, Katrin Horn, Kevin Ho, Lars Wallentin, Man Li, Markus Scholz, Mathias Gorski, Matthias Wuttke, Myriam Rheinberger, Niek Verweij, Owen M. Woodward, Pavel Hamet, Pim van der Harst, Sahar Ghasemi, Sanaz Sedaghat, Sarah A. Pendergrass, Shih-Jen Hwang, Veronique Vitart, Victoria L. Halperin Kuhns, Wei Huang, Wolfgang Koenig, Yizhe Xu, Yong Li

Genotyping: Alan B. Zonderman, Alexander Teumer, André G. Uitterlinden, Antje Körner, Archie Campbell, Ayse Demirkan, Blair H. Smith, Brenda W.J.H. Penninx, Caroline Hayward, Carsten A. Böger, Cassandra N. Spracklen, Christian Fuchsberger, Cornelia M. van Duijn, Daniela Baptista, Daniela Ruggiero, Daniela Toniolo, David J. Porteous, Dennis O. Mook-Kanamori, Erik Ingelsson, Erika Salvi, Federica Rizzi, Florian Kronenberg, Georg Ehret, Grant W. Montgomery, Harry Campbell, James G. Wilson, Jaspal S. S. Kooner, Johan Ärnlöv, Johanne Tremblay, John C. Chambers, Karen L. Mohlke, Leo-Pekka Lyytikäinen, Leslie A. Lange, Marcus E. Kleber, Melanie Waldenberger, Michael H. Preuss, Michele K. Evans, Michiaki Kubo, Mika Kähönen, Mike A. Nalls, Najaf Amin, Nina Mononen, Olli T. Raitakari, Patrick Sulem, Pavel Hamet, Peter Kovacs, Pim van der Harst, Ralph Burkhardt, Ron T. Gansevoort, Salman M. Tajuddin, Sandosh Padmanabhan, Scott D. Gordon, Simona Vaccargiu, Terho Lehtimäki, Thomas Meitinger, Uwe Völker, Wei Huang, Winfried März, Wolfgang Koenig, Yuri Milaneschi

Functional study: Victoria Halperin Kuhns, Ravchel Lewis, Owen Woodward.

Competing interests

Karsten B. Sieber is full-time employee of GlaxoSmithKline. Gardar Sveinbjornsson and Patrick Sulem are full time employees of deCODE genetics, Amgen Inc. Wolfgang Koenig received modest consultation fees for advisory board meetings from Amgen, DalCor, Kowa, Novartis, Pfizer and Sanofi, and modest personal fees for lectures from Amgen, AstraZeneca, Novartis, Pfizer and Sanofi. Winfried März is employed with Synlab Services GmbH and holds shares of Synlab Holding Deutschland GmbH. Mike A. Nalls is supported by a consulting contract between Data Tecnica International LLC and the National Institute on Aging (NIA), National Institutes of Health (NIH), Bethesda, MD, USA and consults for Illumina Inc., the Michael J. Fox Foundation, and the University of California Healthcare. Lars Wallentin received Institutional grants from GlaxoSmithKline, AstraZeneca, BMS, Boehringer-Ingelheim, Pfizer, MSD and Roche Diagnostics. Harvey White received grants and personal fees from Eli Lilly and Company, grants from National Institute of Health, personal fees from Omthera Pharmaceuticals, Pfizer New Zealand, Elsai Inc., DalCor Pharma UK Inc., Sirtex, Acetelion, from CSL Behring LLC, Luitpold Pharmaceuticals Ltd., Sanofi Aventis, and non-financial support from AstraZeneca, outside the submitted work. Kevin Ho disclosed a research and financial relationship with Sanofi-Genzyme. Bruce M. Psaty serves on the DSMB of a clinical trial funded by the manufacturer (Zoll LifeCor) and on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. Adam S. Butterworth received grants from MSD, Pfizer, Novartis, Biogen and Bioverativ and personal fees from Novartis. Markus Scholz consults for and received grant support from Merck Serono not related to this project. Anna Köttgen received grant support from Gruenenthal not related to this project. Other authors declare no competing interests.

Figure Legends

Figure 1: Trans-ethnic GWAS meta-analysis identifies 183 loci associated with serum urate

Outer ring: Dot size represents the genetic effect size of the index SNP at each labeled locus on serum urate. Blue band: $-\log_{10}(P)$ for association with serum urate, by chromosomal position (GRCh37 (hg19) reference build). Red line indicates genome-wide significance ($P=5\times10^{-8}$). Blue gene labels indicate novel loci, gray labels loci reported in previous GWAS of serum urate. Green band: $-\log_{10}(P)$ for association with gout, by chromosomal position. Red line indicates genome-wide significance ($P=5\times10^{-8}$). Inner band: Dots represent index SNPs with significant heterogeneity and are color-coded according to its source: green for ancestry-related heterogeneity (p-anc-het<2.7×10^{-4} [0.05/183]), red for residual heterogeneity (p-res-het <2.7×10^{-4}), and yellow for both (p-anc-het and p-res-het <2.7×10^{-4}). Loci are labeled with the gene closest to the index SNP.

Figure 2: A genetic risk score (GRS) for serum urate improves gout risk prediction. (A) Histogram of the urate GRS among 334,880 European ancestry participants of the UK Biobank. The Y axes show the number of individuals (left) and the prevalence of gout (right), the X axis shows bins of the urate GRS; **(B)** Y axis displays the age- and sex-adjusted odds ratio of gout by GRS bin (X axis), comparing each other bin to the most prevalent one; **(C)** Comparison of the receiver operating characteristic (ROC) curves of different prediction models of gout: genetic (GRS only; red), demographic (age + sex; green), and combined (GRS + age + sex; blue). Y-axis: sensitivity, X-axis: specificity

Figure 3: Serum urate shows widespread genetic correlations with cardio-metabolic risk factors and diseases.

The Circos plot shows significant genome-wide genetic correlations between serum urate and 214 complex traits or diseases ($p<6.6x10^{-5}$), with bar height proportional to the genetic correlation coefficient (r_g) estimate for each trait and coloring according to its direction (dark blue, $r_g>0$; light blue, $r_g<0$). Traits and diseases are labeled on the outside of the plot, and grouped into nine different categories. Each category is color-coded (inner ring, inset). The greatest genetic correlation was observed with gout ($r_g=0.92$, $p=3.3x10^{-70}$). Genetic correlations with multiple cardio-metabolic risk factors and diseases reflect their known directions from observational studies.

Figure 4: Genes expressed in urate-associated loci are enriched in kidney tissue and pathways. **(A)** Grouped physiological systems (X-axis) that were tested individually for enrichment of expression of genes in urate-associated loci are shown as a bar plot, with the $-\log_{10}$(*P*-value) on the Y-axis. Significantly enriched systems are labeled and highlighted in blue (false discovery rate [FDR] <0.01). **(B)** Correlated (r > 0.2) meta-gene sets that were strongly enriched for genes mapping into urate-associated loci (FDR <0.01). Thickness of the edges represents the magnitude of the correlation coefficient, node size, color and intensity represent the number of clustered gene sets, gene set origin, and enrichment p-value, respectively.

Figure 5: Prioritization of p.Thr139Ile at *HNF4A* and functional study of *HNF4A* regulation of *ABCG2* transcription.

(A) Graph shows credible set size (X-axis) against the posterior probability of association (PPA; Y-axis) for each of 1,453 SNPs with PPA >1% in 114 99% credible sets. Triangles mark missense SNPs, with size proportional to their Combined Annotation Dependent Depletion (CADD) score. Blue triangles indicate missense variants mapping into small (≤5 SNPs) credible sets or with high PPA (≥50%). **(B)** Predicted HNF1A or HNF4A binding sites in the promoter region of *ABCG2*, the consensus affinity sequence, and the p value of likely matches. **(C)** Relative luciferase activity and transactivation of *ABCG2* promoter in cells transfected with variable amount of HNF1A or HNF4A

constructs. ± SD, n=3 independent experiments, * p<0.01. **(D)** Position of p.Thr139Ile (T139I) in DNA binding domain / hinge region within HNF4A homodimer structure (PBD 4IQR). **(E)** Relative luciferase activity and transactivation of *ABCG2* promoter in cells transfected with variable amount of constructs of wild-type HNF4A (threonine) or isoleucine at position 139. ± SD, n=3 independent experiments, * p<0.01.

Figure 6: Co-localization of urate-association signals with gene expression in *cis* in kidney tissues

Serum urate association signals identified among European ancestry individuals were tested for co-localization with all eQTLs where the eQTL *cis*-window overlapped (±100 kb) the index SNP. Genes with ≥1 positive co-localization (posterior probability of one common causal variant, H4, ≥0.80) in a kidney tissue are illustrated with the respective index SNP and transcript (Y-axis). Co-localizations across all tissues (X-axis) are illustrated as dots, where the size of the dots indicates the posterior probability of the co-localization. Negative co-localizations (posterior probability of H4 <0.80) are marked in gray, while the positive co-localizations are color-coded based on the predicted change in expression relative to the allele associated with higher serum urate.

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

References

1. Kraja, A.T. *et al.* New Blood Pressure-Associated Loci Identified in Meta-Analyses of 475 000 Individuals. *Circ Cardiovasc Genet* **10**(2017).
2. Warren, H.R. *et al.* Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet* **49**, 403-415 (2017).
3. Jinno, S., Hasegawa, K., Neogi, T., Goto, T. & Dubreuil, M. Trends in Emergency Department Visits and Charges for Gout in the United States between 2006 and 2012. *J Rheumatol* **43**, 1589-92 (2016).
4. Mikuls, T.R., Farrar, J.T., Bilker, W.B., Fernandes, S. & Saag, K.G. Suboptimal physician adherence to quality indicators for the management of gout and asymptomatic hyperuricaemia: results from the UK General Practice Research Database (GPRD). *Rheumatology (Oxford)* **44**, 1038-42 (2005).
5. Liu, C. *et al.* Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat Genet* **48**, 1162-70 (2016).
6. Yang, Q. *et al.* Genome-wide search for genes affecting serum uric acid levels: the Framingham Heart Study. *Metabolism* **54**, 1435-41 (2005).
7. Becker, M.A. *et al.* Febuxostat compared with allopurinol in patients with hyperuricemia and gout. *N Engl J Med* **353**, 2450-61 (2005).
8. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat Genet* **48**, 1151-1161 (2016).
9. Sveinbjornsson, G. *et al.* Rare mutations associating with serum creatinine and chronic kidney disease. *Hum Mol Genet* **23**, 6935-43 (2014).
10. Li, M. *et al.* SOS2 and ACP1 Loci Identified through Large-Scale Exome Chip Analysis Regulate Kidney Development and Function. *J Am Soc Nephrol* **28**, 981-994 (2017).
11. Tasic, V. *et al.* Clinical and functional characterization of URAT1 variants. *PLoS One* **6**, e28641 (2011).
12. Hagos, Y., Stein, D., Ugele, B., Burckhardt, G. & Bahn, A. Human renal organic anion transporter 4 operates as an asymmetric urate transporter. *J Am Soc Nephrol* **18**, 430-9 (2007).
13. MacArthur, D.G. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469-76 (2014).
14. Xue, Y. *et al.* GPS 2.1: enhanced prediction of kinase-specific phosphorylation sites with an algorithm of motif length selection. *Protein Eng Des Sel* **24**, 255-60 (2011).
15. Diaz-Torne, C., Perez-Herrero, N. & Perez-Ruiz, F. New medications in development for the treatment of hyperuricemia of gout. *Curr Opin Rheumatol* **27**, 164-9 (2015).
16. Kramer, H.J. *et al.* African Ancestry-Specific Alleles and Kidney Disease Risk in Hispanics/Latinos. *J Am Soc Nephrol* **28**, 915-922 (2017).
17. Long, W. *et al.* Critical Roles of Two Hydrophobic Residues within Human Glucose Transporter 9 (hSLC2A9) in Substrate Selectivity and Urate Transport. *J Biol Chem* **290**, 15292-303 (2015).
18. Tin, A. *et al.* Genome-wide association study for serum urate concentrations and gout among African Americans identifies genomic risk loci and a novel URAT1 loss-of-function allele. *Hum Mol Genet* **20**, 4056-68 (2011).
19. Woodward, O.M. *et al.* Identification of a urate transporter, ABCG2, with a common functional polymorphism causing gout. *Proc Natl Acad Sci U S A* **106**, 10338-42 (2009).
20. Sautin, Y.Y. & Johnson, R.J. Uric acid: the oxidant-antioxidant paradox. *Nucleosides Nucleotides Nucleic Acids* **27**, 608-19 (2008).

21. Perez-Ruiz, F., Sundy, J.S., Miner, J.N., Cravets, M. & Storgard, C. Lesinurad in combination with allopurinol: results of a phase 2, randomised, double-blind study in patients with gout with an inadequate response to allopurinol. *Ann Rheum Dis* **75**, 1074-80 (2016).
22. Muraoka, S. & Miura, T. Inhibition by uric acid of free radicals that damage biological molecules. *Pharmacol Toxicol* **93**, 284-9 (2003).
23. Tennessen, J.A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64-9 (2012).
24. Morrison, A.C. *et al.* Practical Approaches for Whole-Genome Sequence Analysis of Heart- and Blood-Related Traits. *Am J Hum Genet* **100**, 205-215 (2017).
25. Kamatani, Y. *et al.* Genome-wide association study of hematological and biochemical traits in a Japanese population. *Nat Genet* **42**, 210-5 (2010).
26. Okada, Y. *et al.* Meta-analysis identifies multiple loci associated with kidney function-related traits in east Asian populations. *Nat Genet* **44**, 904-9 (2012).
27. Jordan, K.M. Up-to-date management of gout. *Curr Opin Rheumatol* **24**, 145-51 (2012).
28. Long, W. *et al.* Identification of Key Residues for Urate Specific Transport in Human Glucose Transporter 9 (hSLC2A9). *Sci Rep* **7**, 41167 (2017).
29. Feig, D.I., Kang, D.H. & Johnson, R.J. Uric acid and cardiovascular risk. *N Engl J Med* **359**, 1811-21 (2008).
30. Nomura, N. *et al.* Structure and mechanism of the mammalian fructose transporter GLUT5. *Nature* **526**, 397-401 (2015).
31. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 7-8 (2015).
32. Pao, S.S., Paulsen, I.T. & Saier, M.H., Jr. Major facilitator superfamily. *Microbiol Mol Biol Rev* **62**, 1-34 (1998).
33. Asano, T. *et al.* The role of N-glycosylation of GLUT1 for glucose transport activity. *J Biol Chem* **266**, 24632-6 (1991).
34. Boyle, E.A., Li, Y.I. & Pritchard, J.K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186 (2017).
35. Chandra, V. *et al.* Multidomain integration in the structure of the HNF-4alpha nuclear receptor complex. *Nature* **495**, 394-8 (2013).
36. Sulem, P. *et al.* Identification of low-frequency variants associated with gout and serum uric acid levels. *Nat Genet* **43**, 1127-30 (2011).
37. Togawa, N., Miyaji, T., Izawa, S., Omote, H. & Moriyama, Y. A Na+-phosphate cotransporter homologue (SLC17A4 protein) is an intestinal organic anion exporter. *Am J Physiol Cell Physiol* **302**, C1652-60 (2012).
38. Sekine, T., Miyazaki, H. & Endou, H. Molecular physiology of renal organic anion transporters. *Am J Physiol Renal Physiol* **290**, F251-61 (2006).
39. Clemencon, B. *et al.* Expression, purification, and structural insights for the human uric acid transporter, GLUT9, using the Xenopus laevis oocytes system. *PLoS One* **9**, e108852 (2014).
40. Maiden, M.C., Davis, E.O., Baldwin, S.A., Moore, D.C. & Henderson, P.J. Mammalian and bacterial sugar transport proteins are homologous. *Nature* **325**, 641-3 (1987).
41. Merkulova, M. *et al.* Targeted deletion of the Ncoa7 gene results in incomplete distal renal tubular acidosis in mice. *Am J Physiol Renal Physiol* **315**, F173-F185 (2018).
42. Kirby, A. *et al.* Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in MUC1 missed by massively parallel sequencing. *Nat Genet* **45**, 299-303 (2013).
43. Dinour, D. *et al.* URAT1 mutations cause renal hypouricemia type 1 in Iraqi Jews. *Nephrol Dial Transplant* **26**, 2175-81 (2011).

44. Nigam, S.K. *et al.* The organic anion transporter (OAT) family: a systems biology perspective. *Physiol Rev* **95**, 83-123 (2015).
45. Lek, M. *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285-91 (2016).
46. Daigo, K. *et al.* Proteomic analysis of native hepatocyte nuclear factor-4alpha (HNF4alpha) isoforms, phosphorylation status, and interactive cofactors. *J Biol Chem* **286**, 674-86 (2011).
47. Matsuo, H. *et al.* ABCG2 dysfunction causes hyperuricemia due to both renal urate underexcretion and renal urate overload. *Sci Rep* **4**, 3755 (2014).
48. Zhu, Q. *et al.* T130I mutation in HNF-4alpha gene is a loss-of-function mutation in hepatocytes and is associated with late-onset Type 2 diabetes mellitus in Japanese subjects. *Diabetologia* **46**, 567-73 (2003).
49. Auer, P.L. *et al.* Rare and Coding Region Genetic Variants Associated With Risk of Ischemic Stroke: The NHLBI Exome Sequence Project. *JAMA Neurol* **72**, 781-8 (2015).
50. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216-20 (2013).
51. Ketharnathan, S. *et al.* A non-coding genetic variant maximally associated with serum urate levels is functionally linked to HNF4A-dependent PDZK1 expression. *Hum Mol Genet* **27**, 3964-3973 (2018).
52. Guerriero, C.J. & Brodsky, J.L. The delicate balance between secreted protein folding and endoplasmic reticulum-associated degradation in human physiology. *Physiol Rev* **92**, 537-76 (2012).
53. Silva, C.T. *et al.* A Combined Linkage and Exome Sequencing Analysis for Electrocardiogram Parameters in the Erasmus Rucphen Family Study. *Front Genet* **7**, 190 (2016).
54. Devlin, B., Roeder, K. & Wasserman, L. Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol* **60**, 155-66 (2001).
55. Shakhbazov, K. *et al.* Shared genetic control of expression and methylation in peripheral blood. *BMC Genomics* **17**, 278 (2016).
56. Marioni, R.E. *et al.* DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol* **16**, 25 (2015).
57. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034-50 (2005).
58. Peters, M.J. *et al.* The transcriptional landscape of age in human peripheral blood. *Nat Commun* **6**, 8570 (2015).
59. Cooper, G.M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* **15**, 901-13 (2005).
60. Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
61. Ziller, M.J. *et al.* Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet* **7**, e1002389 (2011).
62. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
63. Shah, S. *et al.* Improving Phenotypic Prediction by Combining Genetic and Epigenetic Associations. *Am J Hum Genet* **97**, 75-85 (2015).
64. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat Rev Genet* **11**, 499-511 (2010).
65. Illingworth, R.S. & Bird, A.P. CpG islands--'a rough guide'. *FEBS Lett* **583**, 1713-20 (2009).
66. Ziller, M.J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477-81 (2013).

67. Wilk, J.B. *et al.* Segregation analysis of serum uric acid in the NHLBI Family Heart Study. *Hum Genet* **106**, 355-9 (2000).
68. Kuo, C.F. *et al.* Familial aggregation of gout and relative genetic and environmental contributions: a nationwide population study in Taiwan. *Ann Rheum Dis* **74**, 369-74 (2015).
69. Krishnan, E., Lessov-Schlaggar, C.N., Krasnow, R.E. & Swan, G.E. Nature versus nurture in gout: a twin study. *Am J Med* **125**, 499-504 (2012).
70. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
71. Rao, D.C. *et al.* The Clinical Lipid Research Clinic Family Study: familial determinants of plasma uric acid. *Hum Genet* **60**, 257-61 (1982).
72. Emmerson, B.T., Nagel, S.L., Duffy, D.L. & Martin, N.G. Genetic control of the renal clearance of urate: a study of twins. *Ann Rheum Dis* **51**, 375-7 (1992).
73. Hadfield, J. MCMC methods for multi-response generalized linear mixed models: the MCMC glmm R Package. *J Stat Softw* **33**, 1-22 (2010).
74. Pattaro, C. *et al.* The Cooperative Health Research in South Tyrol (CHRIS) study: rationale, objectives, and preliminary results. *J Transl Med* **13**, 348 (2015).
75. Noce, D. *et al.* Sequential recruitment of study participants may inflate genetic heritability estimates. *Hum Genet* **136**, 743-757 (2017).
76. Rice, T. *et al.* Heterogeneity in the familial aggregation of fasting serum uric acid level in five North American populations: the Lipid Research Clinics Family Study. *Am J Med Genet* **36**, 219-25 (1990).
77. Sudlow, C. *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* **12**, e1001779 (2015).
78. Saag, K.G. *et al.* Impact of Febuxostat on Renal Function in Gout Patients With Moderate-to-Severe Renal Impairment. *Arthritis Rheumatol* **68**, 2035-43 (2016).
79. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).
80. Dalbeth, N. *et al.* Effects of Febuxostat in Early Gout: A Randomized, Double-Blind, Placebo-Controlled Study. *Arthritis Rheumatol* **69**, 2386-2395 (2017).
81. Bodenhofer, U., Kothmeier, A. & Hochreiter, S. APCluster: an R package for affinity propagation clustering. *Bioinformatics* **27**, 2463-4 (2011).
82. Goldfarb, D.S., MacDonald, P.A., Gunawardhana, L., Chefo, S. & McLean, L. Randomized controlled trial of febuxostat versus allopurinol or placebo in individuals with higher urinary uric acid excretion and calcium stones. *Clin J Am Soc Nephrol* **8**, 1960-7 (2013).
83. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
84. Sheffield, N.C. *et al.* Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res* **23**, 777-88 (2013).
85. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-30 (2015).
86. Whelton, A., MacDonald, P.A., Chefo, S. & Gunawardhana, L. Preservation of renal function during gout treatment with febuxostat: a quantitative study. *Postgrad Med* **125**, 106-14 (2013).
87. Becker, M.A., MacDonald, P.A., Hunt, B. & Gunawardhana, L. Treating hyperuricemia of gout: safety and efficacy of febuxostat and allopurinol in older versus younger subjects. *Nucleosides Nucleotides Nucleic Acids* **30**, 1011-7 (2011).
88. Goldfarb, D.S., MacDonald, P.A., Hunt, B. & Gunawardhana, L. Febuxostat in gout: serum urate response in uric acid overproducers and underexcretors. *J Rheumatol* **38**, 1385-9 (2011).

89. Gillies, C.E. *et al.* An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome. *Am J Hum Genet* **103**, 232-244 (2018).
90. GTEx. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
91. Augustin, R. The protein family of glucose transport facilitators: It's not only about glucose after all. *IUBMB Life* **62**, 315-33 (2010).
92. Miyata, H. *et al.* Identification of Febuxostat as a New Strong ABCG2 Inhibitor: Potential Applications and Risks in Clinical Situations. *Front Pharmacol* **7**, 518 (2016).
93. Fehrmann, R.S. *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet* **7**, e1002197 (2011).
94. Tan, P.K., Ostertag, T.M. & Miner, J.N. Mechanism of high affinity inhibition of the human urate transporter URAT1. *Sci Rep* **6**, 34995 (2016).
95. Whelton, A., Macdonald, P.A., Zhao, L., Hunt, B. & Gunawardhana, L. Renal function in gout: long-term treatment effects of febuxostat. *J Clin Rheumatol* **17**, 7-13 (2011).
96. Kirsten, H. *et al.* Dissecting the genetics of the human transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding locidagger. *Hum Mol Genet* **24**, 4746-63 (2015).
97. Ichida, K. *et al.* Decreased extra-renal urate excretion is a common cause of hyperuricemia. *Nat Commun* **3**, 764 (2012).
98. Qu, L.H., Jiang, H. & Chen, J.H. Effect of uric acid-lowering therapy on blood pressure: systematic review and meta-analysis. *Ann Med* **49**, 142-156 (2017).
99. Nuki, G., Doherty, M. & Richette, P. Current management of gout: practical messages from 2016 EULAR guidelines. *Pol Arch Intern Med* **127**, 267-277 (2017).
100. Dalbeth, N. *et al.* Discordant American College of Physicians and international rheumatology guidelines for gout management: consensus statement of the Gout, Hyperuricemia and Crystal-Associated Disease Network (G-CAN). *Nat Rev Rheumatol* **13**, 561-568 (2017).
101. Dalbeth, N., Choi, H.K. & Terkeltaub, R. Review: Gout: A Roadmap to Approaches for Improving Global Outcomes. *Arthritis Rheumatol* **69**, 22-34 (2017).
102. Lee, C. & Huang, C.H. LASAGNA-Search 2.0: integrated transcription factor binding site search and visualization in a browser. *Bioinformatics* **30**, 1923-5 (2014).
103. Vesuna, F., Winnard, P., Jr. & Raman, V. Enhanced green fluorescent protein as an alternative control reporter to Renilla luciferase. *Anal Biochem* **342**, 345-7 (2005).