University of Dundee

Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels

Tin, Adrienne; Marten, Jonathan; Halperin Kuhns, Victoria L.; Li, Yong; Wuttke, Matthias

Published in:
Nature Genetics

DOI:
10.1038/s41588-019-0504-x

Publication date:
2019

Document Version
Peer reviewed version

Link to publication in Discovery Research Portal

Citation for published version (APA):
, Tin, A., Marten, J., Halperin Kuhns, V. L., Li, Y., Wuttke, M., Kirsten, H., Sieber, K. B., Qiu, C., Gorski, M., Yu, Z., Giri, A., Sveinbjornsson, G., Li, M., Chu, A. Y., Hoppmann, A., O'Connor, L. J., Prins, B., Nutile, T., ... Köttgen, A. (2019). Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nature Genetics*, *51*(10), 1459–1474 . https://doi.org/10.1038/s41588-019-0504-x

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

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

Download date: 21. Apr. 2025

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], Luke J. O'Connor[18], Bram Prins[19], Teresa Nutile[20], Damia Noce[21], Masato Akiyama[22,23], Massimiliano Cocca[24], Sahar Ghasemi[25,26], Peter J. van der Most[27], Katrin Horn[6,7], Yizhe Xu[16], Christian Fuchsberger[21], Sanaz Sedaghat[28], Saima Afaq[29,30], Najaf Amin[28], Johan Ärnlöv[31,32], Stephan J.L. Bakker[33], Nisha Bansal[34,35], Daniela Baptista[36], Sven Bergmann[37,38,39], Mary L. Biggs[40,41], Ginevra Biino[42], Eric Boerwinkle[43], Erwin P. Bottinger[44,45], Thibaud S. Boutin[3], Marco Brumat[46], Ralph Burkhardt[7,47,48], Eric Campana[46], Archie Campbell[49], Harry Campbell[50], Robert J. Carroll[51], Eulalia Catamo[24], John C. Chambers[52,53,54,55,56], Marina Ciullo[20,57], Maria Pina Concas[24], Josef Coresh[1], Tanguy Corre[37,38,58], Daniele Cusi[59,60], Sala Cinzia Felicita[61], Martin H. de Borst[33], Alessandro De Grandi[21], Renée de Mutsert[62], Aiko P.J. de Vries[63], Graciela Delgado[64], Ayse Demirkan[28], Olivier Devuyst[65], Katalin Dittrich[66,67], Kai-Uwe Eckardt[68,69], Georg Ehret[36], Karlhans Endlich[26,70], Michele K. Evans[71], Ron T. Gansevoort[33], Paolo Gasparini[24,46], Vilmantas Giedraitis[72], Christian Gieger[73,74,75], Giorgia Girotto[24,46], Martin Gögele[21], Scott D. Gordon[76], Daniel F. Gudbjartsson[15], Vilmundur Gudnason[77,78], Toomas Haller[79], Pavel Hamet[80,81], Tamara B. Harris[82], Caroline Hayward[3], Andrew A. Hicks[21], Edith Hofer[83,84], Hilma Holm[15], Wei Huang[85,86], Nina Hutri-Kähönen[87,88], Shih-Jen Hwang[89,90], M. Arfan Ikram[28], Raychel M. Lewis[91], Erik Ingelsson[92,93,94,95], Johanna Jakobsdottir[77,96], Ingileif Jonsdottir[15], Helgi Jonsson[97,98], Peter K. Joshi[50], Navya Shilpa Josyula[99], Bettina Jung[10], Mika Kähönen[100,101], Yoichiro Kamatani[22,102], Masahiro Kanai[22,103], Shona M. Kerr[3], Wieland Kiess[7,66,67], Marcus E. Kleber[64], Wolfgang Koenig[104,105,106], Jaspal S. Kooner[54,55,56,107], Antje Körner[7,66,67], Peter Kovacs[108], Bernhard K. Krämer[64], Florian Kronenberg[109], Michiaki Kubo[110], Brigitte Kühnel[73], Martina La Bianca[24], Leslie A. Lange[111], Benjamin Lehne[29], Terho Lehtimäki[112,113], Lifelines Cohort Study[114], Jun Liu[28], Markus Loeffler[6,7], Ruth J.F. Loos[44,115], Leo-Pekka Lyytikäinen[112,113], Reedik Magi[79], Anubha Mahajan[116,117], Nicholas G. Martin[76], Winfried März[64,118,119], Deborah Mascalzoni[21], Koichi Matsuda[120], Christa Meisinger[121,122], Thomas Meitinger[105,123,124], Andres Metspalu[79], Yuri Milaneschi[125], Million Veteran Program[126], Christopher J. O'Donnell[127,128], Otis D. Wilson[129], J. Michael Gaziano[130], Pashupati P. Mishra[131], Karen L. Mohlke[132], Nina Mononen[112,131], Grant W. Montgomery[133], Dennis O. Mook-Kanamori[62,134], Martina Müller-Nurasyid[105,135,136,137], Girish N. Nadkarni[44,138], Mike A. Nalls[139,140], Matthias Nauck[26,141], Kjell Nikus[142,143], Boting Ning[144], Ilja M. Nolte[27], Raymond Noordam[145], Jeffrey OConnell[146], Isleifur Olafsson[147], Sandosh Padmanabhan[148], Brenda W.J.H. Penninx[125], Thomas Perls[149], Annette Peters[74,75,105], Mario Pirastu[150], Nicola Pirastu[50], Giorgio Pistis[151], Ozren Polasek[152,153], Belen Ponte[154,155], David J. Porteous[49,156], Tanja Poulain[7], Michael H. Preuss[44], Ton J. Rabelink[63,157], Laura M. Raffield[132], Olli T. Raitakari[158,159], Rainer Rettig[160], Myriam Rheinberger[10], Kenneth M. Rice[41], Federica Rizzi[161,162], Antonietta Robino[24], Igor Rudan[50], Alena Krajcoviechova[163,164], Renata Cifkova[163,164], Rico Rueedi[37,38], Daniela Ruggiero[20,57], Kathleen A. Ryan[165], Yasaman Saba[166], Erika Salvi[161,167], Helena Schmidt[168], Reinhold Schmidt[83], Christian M. Shaffer[51], Albert V. Smith[78], Blair H. Smith[169], Cassandra N. Spracklen[132], Konstantin Strauch[135,136], Michael Stumvoll[170], Patrick Sulem[15], Salman M. Tajuddin[71], Andrej Teren[7,171], Joachim Thiery[7,47], Chris H. L. Thio[27], Unnur Thorsteinsdottir[15], Daniela Toniolo[61], Anke Tönjes[172], Johanne Tremblay[80,173], André G. Uitterlinden[174], Simona Vaccargiu[150], Pim van der Harst[175,176,177], Cornelia M. van Duijn[28], Niek Verweij[175], Uwe Völker[26,178], Peter Vollenweider[179], Gerard Waeber[179], Melanie Waldenberger[73,74,105], John B. Whitfield[76], Sarah H. Wild[180], James F. Wilson[3,50], Qiong Yang[144], Weihua Zhang[52,54], Alan B. Zonderman[71], Murielle Bochud[58], James G. Wilson[181], Sarah A. Pendergrass[182], Kevin Ho[183,184], Afshin Parsa[185,186], Peter P. Pramstaller[21], Bruce M. Psaty[187,188], Carsten A. Böger[10,189], Harold Snieder[27], Adam S. Butterworth[190], Yukinori Okada[191,192], Todd L. Edwards[193,194], Kari Stefansson[15], Katalin Susztak[9], Markus Scholz[6,7], Iris M.

Heid[11], Adriana M. Hung**[129,194], Alexander Teumer**[25,26], Cristian Pattaro**[21], 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, PhD MS
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

92 14 Biomedical Laboratory Research and Development, Tennessee Valley Healthcare System
93 (626)/Vanderbilt University, Nashville, TN, USA
94 15 deCODE Genetics, Amgen Inc., Reykjavik, Iceland
95 16 Department of Medicine, Division of Nephrology and Hypertension, University of Utah, Salt Lake City,
96 USA
97 17 Genetics, Merck & Co., Inc., Kenilworth, New Jersey, USA
98 18 Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA
99 19 Strangeways Research Laboratory, University of Cambridge, 2 Worts' Causeway, Cambridge, CB1
100 8RN, UK
101 20 Institute of Genetics and Biophysics Adriano Buzzati-Traverso - CNR, Naples, Italy
102 21 Eurac Research, Institute for Biomedicine (affiliated to the University of Lübeck), Bolzano, Italy
103 22 Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences (IMS), Yokohama
104 (Kanagawa), Japan
105 23 Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka,
106 Japan
107 24 Institute for Maternal and Child Health - IRCCS Burlo Garofolo, Trieste, Italy
108 25 Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
109 26 DZHK (German Center for Cardiovascular Research), Partner Site Greifswald, Greifswald, Germany
110 27 Department of Epidemiology, University of Groningen, University Medical Center Groningen,
111 Groningen, The Netherlands
112 28 Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The
113 Netherlands
114 29 Department of Epidemiology and Biostatistics, Faculty of Medicine, School of Public Health, Imperial
115 College London, London, UK
116 30 Institute of Public health & social sciences, Khyber Medical University, Pakistan
117 31 Department of Neurobiology, Care Sciences and Society, Division of Family Medicine and Primary
118 Care, Karolinska Institutet, Stockholm, Sweden
119 32 School of Health and Social Studies, Dalarna University, Sweden
120 33 Department of Internal Medicine, Division of Nephrology, University of Groningen, University
121 Medical Center Groningen, Groningen, The Netherlands
122 34 Division of Nephrology, University of Washington, Seattle, Washington, USA
123 35 Kidney Research Institute, University of Washington, Seattle, Washington, USA
124 36 Cardiology, Geneva University Hospitals, Geneva, Switzerland
125 37 Department of Computational Biology, University of Lausanne, Lausanne, Switzerland
126 38 Swiss Institute of Bioinformatics, Lausanne, Switzerland
127 39 Department of Integrative Biomedical Sciences, University of Cape Town, Cape Town, South Africa
128 40 Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle,
129 Washington, USA
130 41 Department of Biostatistics, University of Washington, Seattle, Washington, USA
131 42 Institute of Molecular Genetics, National Research Council of Italy, Pavia, Italy
132 43 Human Genetics Centre, University of Texas Health Science Centre, Houston, Texas, USA
133 44 The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai,
134 New York, New York, USA
135 45 Digital Health Centre, Hasso Plattner Institute and University of Potsdam, Potsdam, Germany
136 46 University of Trieste, Department of Medicine, Surgery and Health Sciences, Trieste, Italy
137 47 Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University of Leipzig,
138 Leipzig, Germany

139 48 Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg,
140 Germany
141 49 Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine,
142 University of Edinburgh, Edinburgh, UK
143 50 Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics,
144 University of Edinburgh, Edinburgh, UK
145 51 Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville (Tennessee),
146 USA
147 52 Department of Epidemiology and Biostatistics, Faculty of Medicine, School of Public Health, Imperial
148 113 College London, London, UK
149 53 Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 308232, Singapore
150 54 Department of Cardiology, Ealing Hospital, Middlesex UB1 3HW, UK
151 55 Imperial College Healthcare NHS Trust, Imperial College London, London, UK
152 56 MRC-PHE Centre for Environment and Health, 323 School of Public Health, Imperial College London,
153 London, UK
154 57 IRCCS Neuromed, Pozzilli, Italy
155 58 Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland
156 59 Institute of Biomedical Technologies, Italy National Research Council, Segrate (Milano), Italy
157 60 Bio4Dreams - business nursery for life sciences, Bresso (Milano), Italy
158 61 San Raffaele Research Institute, Milano, Italy
159 62 Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, The Netherlands
160 63 Section of Nephrology, Department of Internal Medicine, Leiden University Medical Centre, Leiden,
161 The Netherlands
162 64 5th Department of Medicine (Nephrology, Hypertensiology, Rheumatology, Endocrinology,
163 Diabetology), Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
164 65 Institute of Physiology, University of Zurich, Zurich, Switzerland
165 66 Department of Women and Child Health, Hospital for Children and Adolescents, University of Leipzig,
166 Leipzig, Germany
167 67 Centre for Pediatric Research, University of Leipzig, Leipzig, Germany
168 68 Intensive Care Medicine, Charité, Berlin, Germany
169 69 Department of Nephrology and Hypertension, Friedrich-Alexander-University Erlangen-Nürnberg
170 (FAU), Germany
171 70 Department of Anatomy and Cell Biology, University Medicine Greifswald, Greifswald, Germany
172 71 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural
173 Research Program, National Institutes of Health, Baltimore, Maryland, USA
174 72 Department of Public Health and Caring Sciences, Molecular Geriatrics, Uppsala University, Uppsala,
175 Sweden
176 73 Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Centre
177 for Environmental Health, Neuherberg, Germany
178 74 Institute of Epidemiology, Helmholtz Zentrum München - German Research Centre for Environmental
179 Health, Neuherberg, Germany
180 75 German Center for Diabetes Research (DZD), Neuherberg, Germany
181 76 QIMR Berghofer Medical Research Institute, Brisbane, Australia
182 77 Icelandic Heart Association, Kopavogur, Iceland
183 78 Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland
184 79 Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
185 80 Montreal University Hospital Research Centre, CHUM, Montreal, Canada

186 81 Medpharmgene, Montreal, Canada
187 82 Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural
188 Research Program, National Institutes of Health, Bethesda, Maryland, USA
189 83 Clinical Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz,
190 Austria
191 84 Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz,
192 Austria
193 85 Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese
194 National Human Genome Centre, Shanghai, China
195 86 Shanghai Industrial Technology Institute, Shanghai, China
196 87 Department of Pediatrics, Tampere University Hospital, Tampere, Finland
197 88 Department of Pediatrics, Faculty of Medicine and Life Sciences, University of Tampere, Finland
198 89 NHLBIs Framingham Heart Study, Framingham (Massachusetts), USA
199 90 The Centre for Population Studies, NHLBI, Framingham (Massachusetts), USA
200 91 Department of Physiology, University of Maryland School of Medicine
201 92 Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of
202 Medicine, Stanford, USA
203 93 Stanford Cardiovascular Institute, Stanford University, USA
204 94 Molecular Epidemiology and Science for Life Laboratory, Department of Medical Sciences, Uppsala
205 University, Uppsala, Sweden
206 95 Stanford Diabetes Research Center, Stanford University, Stanford, USAs
207 96 The Centre of Public Health Sciences, University of Iceland, Reykjavik, Iceland
208 97 Landspitalinn University Hospital, Iceland
209 98 University of Iceland, Iceland
210 99 Geisinger Research, Biomedical and Translational Informatics Institute, Rockville, USA
211 100 Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland
212 101 Department of Clinical Physiology, Finnish Cardiovascular Research Center - Tampere, Faculty of
213 Medicine and Health Technology, Tampere University, Tampere, Finland
214 102 Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate
215 School of Medicine, Kyoto, Japan
216 103 Department of Biomedical Informatics, Harvard Medical School, Boston, USA
217 104 Deutsches Herzzentrum München, Technische Universität München, Munich, Germany
218 105 DZHK (German Centre for Cardiovascular Research), Partner Site Munich Heart Alliance, Munich,
219 Germany
220 106 Institute of Epidemiology and Biostatistics, University of Ulm, Ulm, Germany
221 107 National Heart and Lung Institute, Imperial College London, London W12 0NN, UK
222 108 Integrated Research and Treatment Centre Adiposity Diseases, University of Leipzig, Leipzig,
223 Germany
224 109 Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical
225 Pharmacology, Medical University of Innsbruck, Innsbruck, Austria
226 110 RIKEN Centre for Integrative Medical Sciences (IMS), Yokohama (Kanagawa), Japan
227 111 Division of Biomedical Informatics and Personalized Medicine, School of Medicine, University of
228 Colorado Denver - Anschutz Medical Campus, Aurora (Colorado), USA
229 112 Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland
230 113 Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of
231 Medicine and Life Sciences, Tampere University, Tampere, Finland
232 114 Lifelines Cohort Study, Groningen, the Netherlands

233 115 The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New
234 York, New York, USA
235 116 Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
236 117 Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, UK
237 118 Synlab Academy, Synlab Holding Deutschland GmbH, Mannheim, Germany
238 119 Clinical Institute of Medical and Chemical Laboratory Diagnostics, Medical University of Graz,
239 Austria
240 120 Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of
241 Tokyo, Tokyo, Japan
242 121 Independent Research Group Clinical Epidemiology, Helmholtz Zentrum München, German
243 Research Centre for Environmental Health, Neuherberg, Germany
244 122 Chair of Epidemiology Ludwig- Maximilians-Universität München at UNIKA-T Augsburg, Augsburg,
245 Germany
246 123 Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany
247 124 Institute of Human Genetics, Technische Universität München, Munich, Germany
248 125 Department of Psychiatry, Amsterdam Neuroscience and Amsterdam Public Health Research
249 Institute, Amsterdam University Medical Centers, Amsterdam, The Netherlands
250 126 Department of Veterans Affairs, Office of Research and Development, Washington, DC, USA
251 127 VA Boston Healthcare System, Boston, MA, USA
252 128 Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
253 129 Vanderbilt University Medical Centre, Division of Nephrology & Hypertension, Nashville, TN, USA
254 130 Massachusetts Veterans Epidemiology Research and Information Center, VA Cooperative Studies
255 Program, VA Boston Healthcare System, Boston (Massachusetts), USA
256 131 Department of Clinical Chemistry, Finnish Cardiovascular Research Center - Tampere, Faculty of
257 Medicine and Life Sciences, University of Tampere, Tampere, Finland
258 132 Department of Genetics, University of North Carolina, Chapel Hill (North Carolina), USA
259 133 University of Queensland, St Lucia, Australia
260 134 Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, The
261 Netherlands
262 135 Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Centre for
263 Environmental Health, Neuherberg, Germany
264 136 Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Germany
265 137 Department of Internal Medicine I (Cardiology), Hospital of the Ludwig-Maximilians-University
266 (LMU) Munich, Munich, Germany
267 138 Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New
268 York, New York, USA
269 139 Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda,
270 Maryland, USA
271 140 Data Tecnica International, Glen Echo, Maryland, USA
272 141 Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald,
273 Germany
274 142 Department of Cardiology, Heart Center, Tampere University Hospital, Tampere, Finland
275 143 Department of Cardiology, Finnish Cardiovascular Research Center - Tampere, Faculty of Medicine
276 and Life Sciences, Tampere University, Tampere, Finland
277 144 Department of Biostatistics, Boston University School of Public Health, Boston (Massachusetts), USA
278 145 Section of Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical
279 Centre, Leiden, The Netherlands

280 146 University of Maryland School of Medicine, Baltimore, USA
281 147 Department of Clinical Biochemistry, Landspitali University Hospital, Reykjavik, Iceland
282 148 Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK
283 149 Department of Medicine, Geriatrics Section, Boston Medical Center, Boston University School of
284 Medicine, Boston (Massachusetts), USA
285 150 Institute of Genetic and Biomedical Research, National Research Council of Italy, UOS of Sassari, Li
286 Punti (Sassari), Italy
287 151 Department of Psychiatry, University Hospital of Lausanne, Lausanne, Switzerland
288 152 Faculty of Medicine, University of Split, Split, Croatia
289 153 Gen-info Ltd, Zagreb, Croatia
290 154 Service de Néphrologie, Geneva University Hospitals, Geneva, Switzerland
291 155 Nephrology Service, Department of Specialties in Internal Medicine, University Hospitals of Geneva,
292 Switzerland
293 156 Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK
294 157 Einthoven Laboratory of Experimental Vascular Research, Leiden University Medical Centre, Leiden,
295 The Netherlands
296 158 Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland
297 159 Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku,
298 Finland
299 160 Institute of Physiology, University Medicine Greifswald, Karlsburg, Germany
300 161 Department of Health Sciences, University of Milan, Milano, Italy
301 162 ePhood Scientific Unit, ePhood SRL, Milano, Italy
302 163 Centre for Cardiovascular Prevention, First Faculty of Medicine, Department of Medicine, Charles
303 University in Prague, Prague, Czech Republic
304 164 Thomayer Hospital, Prague, Czech Republic
305 165 Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine,
306 Baltimore, USA
307 166 Molecular Biology and Biochemistry, Gottfried Schatz Research Centre for Cell Signaling,
308 Metabolism and Aging, Medical University of Graz, Graz, Austria
309 167 Neuroalgology Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy
310 168 Institute of Molecular Biology and Biochemistry, Centre for Molecular Medicine, Medical University
311 of Graz, Graz, Austria
312 169 Division of Population Health and Genomics, Ninewells Hospital and Medical School, University of
313 Dundee, Dundee, UK
314 170 Division of Endocrinology, Nephrology and Rheumatology, University of Leipzig, Leipzig, Germany
315 171 Heart Centre Leipzig, Leipzig, Germany
316 172 Department of Endocrinology and Nephrology, University of Leipzig, Leipzig, Germany
317 173 CRCHUM, Montreal, Canada
318 174 Department of Internal Medicine, Erasmus MC, University Medical Center Rotterdam, Rotterdam,
319 The Netherlands
320 175 Department of Cardiology, University of Groningen, University Medical Center Groningen,
321 Groningen, The Netherlands
322 176 Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen,
323 The Netherlands
324 177 Durrer Centre for Cardiovascular Research, The Netherlands Heart Institute, Utrecht, The
325 Netherlands

178 Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany
179 Internal Medicine, Department of Medicine, Lausanne University Hospital, Lausanne, Switzerland
180 Centre for Population Health Sciences, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK
181 Department of Physiology and Biophysics, University of Mississippi Medical Centre, Jackson (Mississippi), USA
182 Geisinger Research, Biomedical and Translational Informatics Institute, Danville, Pennsylvania, USA
183 Kidney Health Research Institute (KHRI), Geisinger, Danville, Pennsylvania, USA
184 Department of Nephrology, Geisinger, Danville, Pennsylvania, USA
185 Division of Kidney, Urologic and Hematologic Diseases, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, USA
186 Department of Medicine, University of Maryland School of Medicine, Baltimore, USA
187 Cardiovascular Health Research Unit, Department of Medicine, Department of Epidemiology, Department of Health Service, University of Washington, Seattle, Washington, USA
188 Kaiser Permanente Washington Health Research Institute, Seattle, Washington, USA
189 Department of Nephrology and Rheumatology, Kliniken Südostbayern AG, Regensburg, Germany
190 Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
191 Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences (IMS), Osaka, Japan
192 Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan
193 Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, Nashville, TN, USA
194 Department of Veterans Affairs, Tennessee Valley Healthcare System (626)/Vanderbilt University, Nashville, TN, USA

355 **Abstract**

356 Elevated serum urate levels cause gout, and correlate with cardio-metabolic diseases via poorly
357 understood mechanisms. We performed a trans-ethnic genome-wide association study of
358 serum urate among 457,690 individuals, identifying 183 loci (147 novel) that improve prediction
359 of gout in an independent cohort of 334,880 individuals. Serum urate showed significant
360 genetic correlations with many cardio-metabolic traits, with genetic causality analyses
361 supporting a substantial role for pleiotropy. Enrichment analysis, fine-mapping of urate-
362 associated loci and co-localization with gene expression in 47 tissues implicated kidney and liver
363 as main target organs and prioritized potentially causal genes and variants, including the
364 transcriptional master regulators in liver and kidney, *HNF1A* and *HNF4A*. Experimental
365 validation showed that *HNF4A* trans-activated the promoter of the major urate transporter
366 *ABCG2* in kidney cells, and that HNF4A p.Thr139Ile is a functional variant. Transcriptional co-
367 regulation within and across organs may be a general mechanism underlying the observed
368 pleiotropy between urate and cardio-metabolic traits.

369

370

371

372

373 **Introduction**

374

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

387 Genetic heritability of serum urate varies between 30% and 60% in diverse populations[6-
388 11]. Candidate gene and early genome-wide associations studies (GWAS) have identified three
389 genes as major determinants of urate levels: *SLC2A9*, *ABCG2*, and *SLC22A12*[7,12-18]. While *SLC2A9*
390 and *ABCG2* harbor common variants of relatively large effect[19], *SLC22A12* contains many rare or
391 low-frequency variants associated with lower serum urate levels[20]. The largest GWAS meta-
392 analyses performed to date identified 28 associated genomic loci among European ancestry
393 (EA) individuals[21] and 27 among Japanese individuals[22]. Many genes in the associated loci
394 encode urate transporters or their regulators in kidney and gut, while others are relevant to
395 glucose and lipid metabolism, central functions of the liver, where uric acid is generated. Earlier
396 GWAS did not perform fine-mapping coupled to functional annotation or co-localization with
397 gene expression across tissues to prioritize target tissues, pathways, and potentially causal
398 genes and variants. These approaches have only recently become available owing to increased
399 public availability of large datasets[23,24].

400 Here, we perform a trans-ethnic GWAS meta-analysis of serum urate among 457,690
401 individuals and identify 183 associated genetic loci that improve risk prediction of gout in an
402 independent sample of 334,880 individuals from the UK Biobank. We evaluate the genetic
403 correlation of serum urate with hundreds of cardio-metabolic traits and diseases, and use a
404 recently developed latent causal variable model to examine the contribution of causality versus
405 pleiotropy. We prioritize target variants, genes, tissues and pathways that contribute to the
406 complex regulation of urate levels through comprehensive data integration. To validate the
407 prioritization workflow, we conduct proof-of-principle experimental studies showing that
408 HNF4A, a transcriptional master regulator in liver and kidney proximal tubule, can regulate
409 transcription of the major urate transporter ABCG2 in kidney cells and that the fine-mapped
410 *HNF4A* variant p.Thr139Ile is functional. Transcriptional co-regulation of processes linked to
411 energy metabolism within and across organs may underlie the pleiotropy we uncovered
412 between urate levels and numerous cardio-metabolic traits.

413

414 **Results**

415 **Meta-analyses and characterization of serum urate-associated loci**

416 ***Overview***

417 Trans-ethnic meta-analyses were conducted to maximize the sample size for studying the
418 genetic landscape of serum urate. EA-specific analyses were used where population-specific
419 linkage disequilibrium (LD) was required for prioritizing urate-associated genes, tissues, and
420 pathways, identifying genetic correlations with other traits, and to perform gout risk prediction
421 (**Supplementary Figure 1**).

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

423 The primary trans-ethnic GWAS meta-analysis included 457,690 individuals (EA, n=288,649;
424 East Asian ancestry [EAS], n=125,725; African Americans [AA], n=33,671; South Asian ancestry
425 [SA], n=9,037; and Hispanics [HIS], n=608) from 74 studies, with mean urate levels ranging from
426 4.2 to 7.2 mg/dl (**Supplementary Table 1**). GWAS were performed based on genotypes imputed

427 using the 1000 Genomes Project or Haplotype Reference Consortium reference panels
428 (Methods, **Supplementary Table 2**). Following standardized study-specific quality control and
429 variant filtering procedures, we combined results through inverse-variance weighted fixed
430 effect meta-analysis. There was no evidence of inflation due to unmodeled population
431 structure (LD score regression intercept=1.01; genomic inflation factor λ_{GC}=1.04). Post-meta-
432 analysis variant filtering left 8,249,849 high-quality SNPs for downstream analyses (Methods).

433 We identified 183 loci, defined as the +/-500 kb region around the SNP with the lowest
434 p-value (index SNP), that contained at least one SNP associated at genome-wide significance
435 ($p \leq 5x10^{-8}$, **Figure 1**, **Supplementary Table 3**). Of these, 36 contained a SNP reported as the
436 index SNP in previous GWAS of serum urate[13,15,17,18,21,22,25,26], and 147 were considered novel
437 (**Figure 1**). Absolute effect estimates of each copy of the respective index SNP on serum urate
438 ranged from 0.28 mg/dl (known *SLC2A9* locus) to 0.017 mg/dl (novel *KLB* locus). The average
439 absolute effect across all index SNPs was of 0.038 mg/dl (standard deviation [SD] 0.033).
440 Regional association plots for all 183 loci are shown in **Supplementary Figure 2**.

441 The index SNPs at all 183 loci explained an estimated 7.7% of the serum urate variance
442 (Methods), as compared to 5.3% of the variance explained by variants previously reported by
443 GWAS in EA populations[21]. In a large participating general population-based pedigree study, the
444 183 index SNPs explained 17% of serum urate genetic heritability (h^2=37%, 95% credible
445 interval: 29%, 45%). The index SNPs at the three major urate loci *SLC2A9*, *ABCG2* and *SLC22A12*
446 explained 5% of the genetic heritability (**Supplementary Figure 3**; Methods).

447 ***Characterization of ancestry-related heterogeneity***

448 For the 183 index SNPs, we observed no evidence of systematic between-study heterogeneity
449 (median I^2=2%, interquartile range 0-14%; **Supplementary Table 3**). Fourteen index SNPs
450 showed significant evidence of ancestry-associated heterogeneity ($p_{anc-het} < 2.7x10^{-4} = 0.05/183$)
451 when tested using meta-regression (**Supplementary Figure 4**, Methods), consistent with their
452 higher measures of between-study heterogeneity ($I^2 > 25\%$, **Figure 1**, **Supplementary Table 3**).
453 The most significant ancestry-associated heterogeneity was observed for the index SNP
454 rs3775947 at *SLC2A9* ($p_{anc-het}=1.5x10^{-127}$, effect per copy of the coded allele in EA 0.34 mg/dl, AA

455 0.26 mg/dl, EAS 0.17 mg/dl, HIS 0.41 mg/dl, and SA 0.21 mg/dl), consistent with previous
456 reports of population heterogeneity of genetic effects at this locus[27]. In addition, nine genome-
457 wide significant loci were identified through meta-regression that did not overlap with the 183
458 significant loci from the primary trans-ethnic fixed-effects meta-analysis. Of these, the index
459 SNPs at *SLC2A2* and *KCNQ1* were genome-wide significant in EAS (**Supplementary Table 4**).
460 Results from ancestry-specific meta-analyses of EA, AA, EAS and SA are summarized in
461 **Supplementary Tables 5 to 8**, respectively, as well as in the **Supplementary Information**.

462 ***Sex-stratified meta-analyses of serum urate GWAS***

463 Mean serum urate levels and gout risk are higher in men than in women[28]. We therefore
464 performed secondary sex-specific analyses to evaluate whether the 183 urate-associated index
465 SNPs showed sex-specific differences. After multiple-testing correction, six SNPs showed
466 significant effect differences ($P_{diff}<2.7x10^{-4}=0.05/183$), at *SLC2A9*, *ABCG2*, *CAPN1*, *GCKR*, *IDH2*,
467 and *SLC22A12* (**Supplementary Table 9**). A genome-wide test for differences in genetic effects
468 on urate levels between men and women identified only SNPs at *SLC2A9* and *ABCG2* as
469 significant ($p_{diff}<5x10^{-8}$, Methods*,* **Supplementary Figure 5**), consistent with previous
470 reports,[7,14,15,21] with few additional loci outside of the extended sex-specific regions that were
471 suggestive of sex differences ($p_{diff}<1x10^{-5}$, **Supplementary Table 10**).
472

473 ***Epidemiological and clinical landscape***

474 *Urate-associated SNPs are associated with gout*

475 To assess the association of the 183 trans-ethnic urate index SNPs with gout, we investigated
476 their effects in a trans-ethnic meta-analysis of gout from 20 studies, based on 763,813
477 participants including 13,179 with gout (Methods, **Figure 1**, **Supplementary Table 1**). Consistent
478 with the causal role of hyperuricemia in gout, genetic effects were highly correlated (Spearman
479 correlation coefficient 0.87, **Supplementary Figure 6A**); 55 SNPs were significantly associated
480 with gout ($p<2.7x10^{-4}=0.05/183$). In agreement with previous findings[29], the largest odds ratio
481 (OR) for gout was observed at *ABCG2* (rs74904971, OR 2.04, 95% confidence interval [CI] 1.96-

482 2.12, $P=7.7 \times 10^{-299}$). The genetic effects were generally higher among index SNPs with lower
483 minor allele frequency (MAF), with the exception of a few large-effect SNPs with MAF>10%,
484 mapping into loci that encode urate transporters with known major effects on urate levels:
485 *SLC2A9*, *ABCG2*, and *SLC22A12*[30] (**Supplementary Figure 6B**).
486
487 *A genetic risk score for urate improves risk prediction for gout*

488 We evaluated whether a weighted urate genetic risk score (GRS) improved risk prediction of
489 gout when added to demographic information in a large, independent sample of 334,880
490 individuals from the UK Biobank (UKBB), including 4,908 gout cases (Methods). Across
491 categories of the urate GRS, gout prevalence increased from 0.1% in the lowest GRS category to
492 12.9% in the highest GRS category (**Figure 2A, Supplementary Table 11**). Using the most
493 common GRS category as the reference, the age- and sex-adjusted OR of gout ranged from 0.09
494 (95% CI 0.02-0.37, $p=7.8 \times 10^{-4}$) in the lowest GRS category to 13.6 (95% CI 7.2-25.7, $p=1.4 \times 10^{-15}$)
495 in the highest GRS category (**Figure 2B, Supplementary Table 11**). Of note, the 3.5% of
496 individuals in the three highest GRS categories had a >3-fold increase in the risk of gout
497 compared to individuals in the most common GRS category. This risk is comparable to a
498 monogenic disease of modest effect size[31], but affects a comparatively higher proportion of the
499 population.
500 We additionally constructed gout risk prediction models in the independent UK Biobank
501 sample by regressing gout status on the GRS alone ("genetic model"), on age and sex
502 ("demographic model"), and on the GRS, age, and sex ("combined model") in a random training
503 subset consisting of 90% of the individuals. These models were then used to predict gout status
504 in the remaining 10%. The genetic model (area under the receiver operating characteristic
505 curve [AUC]=0.68) was a weaker predictor than the demographic model (AUC=0.79), but
506 addition of the GRS to the demographic model (combined model) significantly increased the
507 prediction accuracy (AUC=0.84, DeLong's test $p<2.2 \times 10^{-16}$; **Figure 2C**). The combined model
508 achieved a sensitivity of 84% and specificity of 68% (Methods). The GRS represents a life-long
509 predisposition to higher urate levels and can be calculated at birth without measurement of

510 serum urate. As opposed to synovial fluid analysis or CT-based imaging to diagnose gout, the
511 GRS is less invasive and avoids radiation exposure, Thus, the GRS may have utility to identify
512 individuals with a high genetic predisposition for gout, allowing for compensatory lifestyle
513 choices to be made earlier in life to reduce the risk of developing gout.
514

515 *High genetic correlations of serum urate with cardio-metabolic traits*

516 Serum urate is positively correlated with many cardio-metabolic risk factors and diseases[32]. We
517 assessed genetic correlations between serum urate and 748 complex traits using cross-trait LD
518 score regression (Methods). Serum urate levels were significantly ($p<6.6x10^{-5}$=0.05/748)
519 genetically correlated with 214 complex traits and diseases (**Supplementary Table 12**). The
520 highest positive genetic correlation (r_g) was with gout (r_g=0.92, p=$3.3x10^{-70}$), followed by traits
521 representing components of the metabolic syndrome such as HOMA-IR (r_g=0.49) and fasting
522 insulin (r_g=0.45). Significant positive genetic correlations were also observed for other cardio-
523 metabolic traits or diseases, including waist circumference, obesity, and type 2 diabetes (**Figure
524 3**). The largest negative correlations were observed with HDL cholesterol-related
525 measurements (r_g up to -0.46), and with estimated glomerular filtration rate (r_g=-0.38 and -0.26
526 for cystatin C-based and creatinine-based eGFR, respectively), consistent with the known role
527 of the kidneys in urate excretion. Overall, the genetic correlations between serum urate and
528 other complex traits and diseases were consistent with observational associations of serum
529 urate levels with cardio-metabolic traits in epidemiological studies[32].

530 To examine whether these genetic correlations reflect causal relationships or pleiotropy,
531 we applied a recently-developed latent causal variable model to estimate the genetic causality
532 proportion (GCP) for seven commonly-studied cardio-metabolic traits (Methods). As a positive
533 control, we analyzed gout, confirming a genetically causal effect of urate on gout (GCP=0.79,
534 **Supplementary Table 13**). Conversely, we identified a range of GCP values consistent with
535 mostly or partially genetically causal effects of the seven cardio-metabolic traits on serum urate
536 levels. The highest GCP estimates were observed for adiposity-related traits (e.g. GCP=0.84 for
537 waist circumference, **Supplementary Table 13**), consistent with higher cell numbers resulting in

538 higher production of purines and consequently urate, as well as with a Mendelian
539 Randomization study that reported a causal effect of adiposity on urate levels.[33] HDL
540 cholesterol levels, on the other hand, showed smaller GCP estimates (GCP<0.5; **Supplementary**
541 **Table 13**), suggesting the existence of a genetic process with a causal effect on both HDL
542 cholesterol and urate. A potential explanation for such a process are co-regulated metabolic
543 processes in the liver that influence both cholesterol and urate levels. These processes may
544 explain a large fraction of heritability for cholesterol levels and a modest fraction of heritability
545 for urate, a type of asymmetry expected to produce a partially genetically causal relationship
546 consistent with the one observed.

547 ***Identification of enriched tissues and pathways***

548 To identify molecular mechanisms and tissues relevant for urate metabolism and handling, and
549 to provide potential clues to the observed genetic correlation with other traits and diseases, we
550 investigated which tissues, cell types and systems may be significantly enriched for the
551 expression of genes mapping into the urate-associated loci. Based on all SNPs with $P<1x10^{-5}$
552 (Methods), we identified significant enrichment (false discovery rate [FDR] <0.01) for 19
553 physiological system entries, three tissues, and two cell types (**Supplementary Table 14**). The
554 strongest enrichment was observed for kidney ($P=9.5x10^{-9}$) and urinary tract ($P=9.9x10^{-9}$), both
555 within the urogenital system, consistent with the kidney's prominent role in controlling serum
556 urate concentrations. Additional significant enrichments were observed in the endocrine and
557 digestive system, including liver, the major site of urate production. Interestingly, a novel
558 significant enrichment was also observed in the musculoskeletal system, specifically for synovial
559 membrane, joint capsule, and joints (**Figure 4A**), the sites of gout attacks.

560 We next tested for cell-type groups with evidence for enriched heritability based on cell-
561 type specific functional genomic elements using stratified LD score regression (Methods). The
562 strongest heritability enrichment was observed for kidney (11.5-fold), followed by liver (5.39-
563 fold; **Supplementary Table 15**). This approach complemented the gene expression-based
564 approach and also supported kidney and liver as the major organs of urate homeostasis.

565 Lastly, we tested whether any gene sets were enriched for variants associated with
566 urate at $P<10^{-5}$ (Methods). Significant enrichment (FDR <0.01) was observed for 383
567 reconstituted gene sets (**Supplementary Table 16**). Since many of these contained overlapping
568 groups of genes, we used affinity propagation clustering to identify 57 meta gene sets
569 (Methods, **Supplementary Table 17**), including a prominent group of inter-correlated gene sets
570 related to kidney and liver development, morphology and function (**Figure 4B**). Together, these
571 analyses underscore the prominent role of the kidney and liver in regulating serum urate levels
572 and implicate the kidney as a major target organ for lowering of serum urate levels.
573

574 **Prioritization of urate loci based on statistical fine-mapping, functional annotation, and gene**
575 **expression**

576 To prioritize target SNPs and genes for translational research, we established a workflow that
577 combined fine-mapping of urate-associated loci with functional annotation and a systematic
578 evaluation of tissue-specific differential gene expression.

579 *Statistical fine-mapping prioritizes candidate SNPs*

580 To identify independent and potentially causal variants, statistical fine-mapping was performed
581 starting from the 123 genome-wide significant loci identified in the EA-specific meta-analysis,
582 because the workflow included methods that used LD estimates from an ancestry-matched
583 reference panel (Methods)[34]. After combining the 123 loci into 99 larger genomic regions based
584 on LD between index SNPs, stepwise model selection in each region identified 114 independent
585 SNPs (r^2<0.01, Methods). Overall, 87 regions contained one independent signal, ten contained
586 two independent SNPs, the *ABCG2* locus contained three and the *SLC2A9* locus four
587 independent SNPs (**Supplementary Table 18**). For each of these 114 independent SNPs, we
588 computed 99% credible sets representing the smallest set of SNPs which collectively account
589 for 99% posterior probability of containing the variant(s) driving the association signal[35]. The
590 99% credible sets contained a median of 16 SNPs (Q1, Q3: 6, 57), and six of them only a single
591 SNP, mapping in or near *INSR*, *RBM8A*, *MPPED2*, *HNF4A*, *CPT1C*, and *SLC2A9* (**Supplementary**
592 **Table 18**). Among 28 small credible sets (≤5 SNPs), several mapped in or near genes with an

593 established role in regulating urate levels such as *SLC2A9*, *PDZK1*, *ABCG2*, *SLC22A11*, and
594 *SLC16A9*[20]. These credible sets contain the most supported candidate SNPs based on the serum
595 urate association signals and would greatly reduce the number of candidate functional variants
596 to be tested in experimental follow-up studies.

597 To further refine the credible set SNPs, we annotated them with respect to their
598 functional consequence and regulatory potential (Methods). Missense SNPs with posterior
599 probabilities >50% for driving the association signals or mapping into small credible sets were
600 identified in *ABCG2*, *UNC5CL*, *HNF1A*, *HNF4A*, *CPS1*, and *GCKR* (**Supplementary Table 19, Figure**
601 **5A**). All missense SNPs except the one in *GCKR* had a CADD score >15, supporting the SNP and
602 the gene it resides in as potentially causal. Indeed, functional effects have already been
603 demonstrated experimentally for variants rs2231142 (Gln141Lys, r^2=1 to the index SNP
604 rs74904971) in *ABCG2*, rs742493 (p.Arg432Gly) in *UNC5CL*, and rs1260326 (p.Leu446Pro) in
605 *GCKR* (**Table 1**). Non-exonic variants with posterior probabilities of >90% and mapping into
606 open chromatin in enriched tissues were identified in *RBM8A*, *SLC2A9*, *INSR*, *HNF4A, PDZK1*,
607 *NRG4*, *UNC5CL*, and *AAK1* (Methods, **Supplementary Figure 7**, **Supplementary Table 19**). When
608 complemented by evidence of gene expression co-localization, these SNPs may represent
609 causal regulatory variants and highlight their potential effector genes.

610 *Gene prioritization via gene expression co-localization analyses*

611 To systematically assess differential gene expression, we tested for co-localization of the urate
612 association signals with expression quantitative trait loci (eQTL) in *cis* across three kidney tissue
613 resources and 44 GTEx tissues (Methods). High posterior probability of co-localization (H4≥0.8,
614 Methods) supports a trait-associated variant acting through modulation of gene expression in
615 the tissue where co-localization is identified. The eQTLs from the three kidney tissue resources
616 were based on glomerular and tubulo-interstitial portions of micro-dissected kidney biopsies
617 from 187 CKD patients and healthy kidney tissue sections of 96 additional individuals
618 (Methods). We identified co-localization with the expression of 13 genes in at least one kidney
619 tissue (**Figure 6**), the tissue with the strongest enrichment for urate-associated variants.
620 Whereas co-localization of some genes was only observed in kidney (*SLC17A4*, *BICC1*, *UMOD*,

621 *GALNTL5*, *NCOA7*), other genes showed co-localization across several tissues (e.g., *ARL6IP5*).
622 The direction of change in gene expression with higher urate levels could vary for the same
623 gene across tissues. For instance, the allele associated with higher serum urate at the *SLC16A9*
624 locus was associated with higher gene expression in kidney, consistent with a regulatory variant
625 in a transporter mediating the reabsorption of urate. The same allele was associated with lower
626 gene expression in other tissues such as aorta, pointing towards tissue-specific regulatory
627 mechanisms[36]. Details on each of the 13 genes with high posterior probability of a variant
628 underlying the associations with both urate and gene expression in kidney are summarized in
629 **Supplementary Table 20.** Significant co-localizations identified across all 47 tissues
630 (**Supplementary Figure 8**) revealed additional novel insights such as co-localization of the urate
631 association signal with *NFAT5* expression in subcutaneous adipose tissue emphasizing its role in
632 adipogenesis[37], or *PDZK1* expression in colon and ileum, important sites of urate excretion.
633 Lastly, we investigated whether any trans-ethnic index SNPs or their proxies (r^2>0.8)
634 were reproducibly associated with gene expression in *trans* in whole blood or peripheral blood
635 mononuclear cell in several large eQTL studies (**Supplementary Table 21**, Supplementary
636 Information). We identified inter-chromosomal associations between five index SNPs and 16
637 transcripts, that were enriched in the term "cardiovascular disease" based on the Human
638 Disease Ontology database (**Supplementary Information, Supplementary Table 22**).
639

640 ***HNF4A* activates *ABCG2* transcription and HNF4A p.Thr139Ile is a functional variant**
641 The gene and variant prioritization workflow was validated using the identified candidates
642 *HNF1A* and *HNF4A*. Co-regulation of target genes by these master regulators of transcription in
643 kidney proximal tubule cells and liver could potentially explain observed genetic correlations[38].
644 We first tested whether *HNF1A* and *HNF4A* have the potential to affect transcription of
645 the *ABCG2* gene, which encodes for a urate transporter of major importance in humans and
646 represented the locus with the highest risk for gout in our screen. *ABCG2* contains several
647 predicted HNF1A and HNF4A binding sites in its promoter region (**Figure 5B**). A luciferase
648 reporter assay in the human embryonic kidney cell line HEK 293 was used to assess

649 transactivation of the human *ABCG2* promoter by HNF4A and HNF1A proteins (Methods,
650 **Supplementary Figure 9A**). Co-expression of HNF4A significantly increased the *ABCG2*
651 promoter-driven luciferase activity, and the activation was dependent on the transfected
652 *HNF4A* expression vector dose and corresponding levels of HNF4A protein (**Figure 5C,**
653 **Supplementary Figure 9B**). As expected, no increase of luciferase activity occurred with the
654 pGL4 vector without the *ABCG2* promoter that was used as a negative control (**Supplementary**
655 **Figure 9D and 9E**). Results for HNF1A indicated that the observed association of this locus with
656 serum urate is unlikely to occur via activation of *ABCG2* in kidney cells (**Figure 5C**), but *HNF1A*
657 has been reported to activate transcription of *PDZK1*, a regulatory protein for several other
658 renal urate transporters[39,40] and also identified in this study.

659 Next, we tested the functional relevance of the prioritized missense p.Thr139Ile allele in
660 HNF4A (NM_178849.2, isoform 1, Methods). Its location within the hinge/ DNA binding domain
661 (**Figure 5D, Supplementary Figure 9F**) supports potentially altered interactions with targeted
662 promoter regions. The isoleucine to threonine substitution at position 139 significantly
663 increased the transactivation of the *ABCG2* promoter and commensurate luciferase activity as
664 compared to the wild-type threonine (**Figure 5E**), without altering HNF4A protein abundance
665 (**Supplementary Figures 9C**). Thus, HNF4A can activate *ABCG2* transcription in a kidney cell line,
666 and HNFA4 p.Thr139Ile is a functional variant. Increased activation of the urate excretory
667 protein ABCG2 by the allele encoding the isoleucine residue should result in lower serum urate
668 levels, consistent with the observed negative association in our GWAS.

669

670

671 **Discussion**

672

673 This large trans-ethnic GWAS meta-analysis of serum urate levels based on 457,690 individuals
674 represents a four-fold increase in sample size over previous studies[21,22,41] and resulted in the
675 identification of 183 urate-associated loci, 147 of which are novel. A genetic urate risk score led
676 to significant improvements of gout risk prediction among 334,880 independent persons, 3.5%
677 of whom had a risk of gout comparable to a Mendelian disease effect size. Genetic correlation

678 and causality analyses confirmed the causal effect of urate on gout, and were consistent with
679 transcriptional co-regulation as a source of pleiotropy in the widespread genetic correlations
680 between serum urate and cardio-metabolic traits. Tissue and cell type-specific enrichment
681 analyses supported kidney and liver, the sites of urate excretion and generation, as key target
682 tissues. Comprehensive fine-mapping and co-localization analyses with gene expression across
683 47 tissues delivered an extensive list of target genes and SNPs for follow-up studies, of which
684 we experimentally confirmed *HNF4A* p.Thr139Ile as a functional allele involved in
685 transcriptional regulation of urate homeostasis.

686 Major challenges of GWAS are to pinpoint causal genes and variants, and to provide
687 actionable insights into disease-relevant mechanisms and pathways in order to improve disease
688 treatment and prevention. This study developed a comprehensive resource of candidate SNPs,
689 genes, tissues and pathways involved in urate metabolism that will enable a wide range of
690 follow-up studies such as our proof-of-principle validation study. Out of the many novel and
691 biologically plausible findings, we highlight three instances in which co-localization of the serum
692 urate and tissue-specific gene expression signals provides new insights into urate metabolism
693 and the prominent role of the renal proximal tubules. First, co-localization helped to prioritize
694 genes in association peaks that previous GWAS could not resolve. For example, the association
695 signal at chromosome 6p22.2 contains the genes encoding four members of the SLC17
696 transporter family (*SLC17A1*, *SLC17A2*, SLC17A3, and *SLC17A4*). Systematic testing of co-
697 localization across genes and tissues supported a variant underlying the urate association signal
698 and differential gene expression only for *SLC17A4* in kidney, with higher expression associated
699 with higher serum urate. Previous experimental studies have implicated SLC17A4 as a urate
700 exporter in intestine[42], and our data support its yet unappreciated role in urate transport in the
701 human kidney. Second, co-localization with gene expression provided insights into transport
702 processes of the proximal tubule, the major site of urate reabsorption and excretion. For
703 example, the gene product of the candidate *ARL6IP5* has been shown to modulate activity of
704 the glutamate transporter SLC1A1[43,44], dysfunction of which causes aminoaciduria[45]; and
705 deletion of the candidate *NCOA7* in mice results in distal renal tubular acidosis[46]. Third, it is
706 noteworthy that co-localization of the urate association signal was observed with differential

707 expression of *MUC1*, *BICC1* and *UMOD* in kidney. Rare mutations in all three genes are known
708 to cause monogenic cystic kidney diseases[47-49], pointing towards a shared mechanism with
709 respect to their association with urate.

710 Another noteworthy finding from this well-powered study was the significant genetic
711 correlations with many cardio-metabolic traits, with directions matching expectation from
712 known observational associations[50]. Many of these traits, like lipid levels and urate, are
713 influenced by liver metabolism. We estimated the genetic causality proportion for these traits,
714 showing that their genetic correlations are partly driven by overlapping or co-regulated
715 metabolic pathways in the liver and not only by a fully causal effect of cholesterol or insulin
716 levels on urate. Likewise, significant genetic correlations with kidney-related traits such as eGFR
717 may reflect shared regulation of processes in the kidney, the major site of urate excretion.
718 Evidence for transcriptional co-regulation, beyond the known importance of *HNF1A* and *HNF4A*
719 in liver and kidney, is supported by the identification of additional urate loci such as *MLXIPL,*
720 *TCF7L2* and *KLF10* that share associations with other metabolic and/or kidney function traits.
721 The observed pleiotropic effects of many urate-associated variants could thus be the potential
722 manifestation of co-regulation of processes that occur within and across tissues relevant to the
723 implicated traits, a mechanism likely to be prevailing in metabolic, but also other traits.

724 In the kidney, nuclear HNF4A is exclusively detected in epithelial cells of the proximal
725 tubule[51], where it has been reported to regulate the expression of SLC2A9 isoform 1[52] and
726 PDZK1[53]. Kidney-specific deletion of *HNF4A* in mice phenocopies Fanconi renotubular
727 syndrome[54]. Detailed kidney tissues transcriptomic analyses support HNF4A to drive a proximal
728 tubule signature cluster of 221 co-expressed genes, including many candidate genes for urate
729 metabolism and transport[51]. In addition to *HNF4A*, *HNF4G*, and *HNF1A*, ten genes in this cluster
730 of co-expressed genes also map into urate-associated loci identified here (*A1CF, CUBN, LRP2,*
731 *PDZK1, SERPINF2, SLC2A9, SLC16A9, SLC17A1, SLC22A12 and SLC47A1*). In addition, our study
732 establishes that HNF4A can also trans-activate transcription of *ABCG2* in a kidney cell line, the
733 key urate secretory transporter in both gut and kidney epithelium[55]. The T allele, encoding the
734 isoleucine substitution at *HNF4A* T139I, showed higher trans-activation of *ABCG2* transcription
735 compared to the wild-type allele, which should result in increased urate secretion and is

736 consistent with the observed association of the T allele with lower serum urate levels. The
737 genetic variant encoding the T139I substitution is located in a region of the HNF4A protein
738 harboring many causative mutations for monogenic maturity onset diabetes of the young
739 (MODY type 1)[56]. Yet, unlike the severe MODY1 missense mutations R127W, D126Y, and
740 R125W,[57] T139I has not been reported to cause MODY1. Instead, it has been reported to
741 increase the risk of type 2 diabetes mellitus, possibly through a liver-specific loss of HNF4A
742 phosphorylation at T139, and to associate with HDL-cholesterol levels[56,58]. These data point to
743 additional complexities when interpreting pleiotropic effects, because there may be several
744 tissue-specific mechanisms by which genetic variants in transcriptional regulators influence
745 metabolic pathways and urate homeostasis.

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

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

763

764

765 **Online Methods**

766 **Overview of GWAS methods**

767 We developed an automated analysis workflow to collect and integrate results from 74 GWAS
768 of serum urate from five ancestry groups participating in the CKDGen Consortium. We used a
769 distributive model for study-specific GWAS with meta-analyses conducted centrally. An analysis
770 plan was circulated to all participating studies accompanied by custom shell and R scripts for
771 phenotype generation (https://github.com/genepi-freiburg/ckdgen-pheno). Study-specific
772 GWAS were conducted after centralized review and approval of the phenotype summary
773 statistics. Study-specific GWAS results were checked using GWAtoolbox[59], including p-value
774 inflation, allele frequency distribution, imputation quality, and completeness of genotypes.
775 Custom scripts were used to compare imputed allele frequencies to those of ancestry-matched
776 reference panels and to visualize variant positions. In addition, quality metrics, including the
777 genomic control factor[60], were compared across studies. The participants of all studies provided
778 written informed consent. Each study had its research protocol approved by the corresponding
779 local ethics committee.

780 **Phenotype definition, genotyping and imputation in participating studies**

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

787 Each study performed genotyping separately and applied study-specific quality filters
788 prior to phasing and imputation (**Supplementary Table 2**). In each study, haplotypes were
789 estimated using MACH[61], ShapeIT[62], Eagle[63], or Beagle[64]. Imputation of genotypes was
790 conducted using reference panels from the Haplotype Reference Consortium (HRC) version
791 1.1[65], 1000 Genomes Project (1000G) phase 3 v5 ALL, or the 1000G phase 1 v3 ALL[66] and

792 ImputeV2[67], minimac3[68], PBWT[69], the Sanger[70], or the Michigan Imputation Server[68]. The
793 imputed genetic dosages were annotated using NCBI b37 (hg19). Each study provided an
794 imputation quality for each variant: ImputeV2 info score, the MACH/minimac RSQ or the
795 SNPTest info score.

796 **Study-specific association analysis**

797 Each study performed ancestry-specific association analysis of serum urate by generating age-
798 and sex-adjusted residuals of serum urate and regressing the residuals on SNP dosage levels,
799 adjusting for study-specific covariates such as study centers and genetic principal components,
800 assuming an additive genetic model. Gout was analyzed as a binary outcome adjusting for age,
801 sex, genetic principal components, and study-specific covariates. Software used for these
802 regression analyses were EPACTS (*q.emmax* for family based studies and *q.linear* otherwise;
803 https://genome.sph.umich.edu/wiki/EPACTS), SNPTest[71], RegScan[72], RVTEST[73], PLINK 1.90[74],
804 Probabel[75], GWAF[76], GEMMA[77], mach2qtl[78] and R. Family-based studies used methods that
805 accounted for relatedness.

806 **Trans-ethnic, ancestry-specific, and sex-stratified meta-analyses**

807 GWAS results from each study were pre-filtered to retain bi-allelic SNPs with imputation quality
808 score >0.6 and minor allele count (MAC) >10 before inclusion into meta-analysis. Fixed effects
809 inverse-variance weighted meta-analysis was performed using METAL[79] with modifications to
810 output higher precision (six decimal places). Genomic control was applied for each study. The
811 genomic inflation factor λ_{GC}[60] was calculated to assess inflation of the test statistics. For each
812 meta-analysis result (trans-ethnic, ancestry-specific, and sex-specific), we excluded SNPs that
813 were present in <50% of the studies and with a total MAC <400. For ancestry-specific meta-
814 analysis, we additionally excluded SNPs with a heterogeneity I^2-statistic[80] >95%. Genome-wide
815 significance was defined as p-value $<5x10^{-8}$. The LD score regression intercept was calculated to
816 assess the evidence for associations driven by population structure[81]. For downstream
817 characterization, 8,249,849 and 8,217,339 autosomal SNPs were retained in the trans-ethnic
818 and European ancestry meta-analysis, respectively. Ancestry-specific meta-analyses were
819 conducted for European ancestry (EA), African Americans (AA), East Asian (EAS) ancestry, and

820 South Asian (SA) ancestry using the same methods and variant filters as the trans-ethnic meta-
821 analysis.

822 Secondary meta-analyses were performed separately in men and women, using the
823 same analytical approaches. To test for significant difference of association between males and
824 females, we used a two-sample t-test $t = \frac{\beta_M - \beta_F}{\sqrt{SE_M{}^2 + SE_F{}^2}}$

825 where β_M and β_F were beta coefficients in males and females, respectively, and SE_M and
826 SE_F were the standard errors among males and females, respectively.
827

828 **Initial determination and annotation of genome-wide significant loci**

829 For each meta-analysis result, SNPs were assigned to loci by selecting the SNP with the lowest
830 genome-wide p-value as the index SNP and defining the adjacent ±500kb-region as the
831 corresponding locus. This procedure was repeated until no genome-wide significant SNP
832 remained. An ancestry-specific locus was defined as a genome-wide significant locus in an
833 ancestry-specific meta-analysis of which the index SNP did not map into within the ±500 kb
834 intervals of any genome-wide significant loci in the trans-ethnic meta-analysis. Index SNPs were
835 annotated using its position and the nearest gene based on hg19, RefSeq genes, and dbSNP147
836 downloaded from ftp://hgdownload.soe.ucsc.edu/mysql/hg19/ on March 23rd, 2017.
837

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

839 The proportion of phenotypic variance explained by index SNPs was calculated as the sum of
840 the variance explained by each index SNP based on this formula: $\beta^2 \left(\frac{2p(1-p)}{var}\right)$, where β is the
841 beta coefficient and p is the MAF of the SNP, and var is the phenotypic variance. For this study,
842 we used the variance of the age- and sex-adjusted residuals of serum urate in EA participants of
843 the ARIC study as the estimate of the phenotypic variance (variance=1.767).

844 Heritability of age- and sex-adjusted urate was estimated using the R package
845 'MCMCglmm'[82] in the Cooperative Health Research In South Tyrol (CHRIS) study,[83] a

846 participating pedigree-based study of EA individuals, comprising 4,373 individuals in 186
847 pedigrees of up to five generations. We estimated overall heritability, heritability excluding
848 index SNPs in the three major urate loci (*SLC2A9*, *ABCG2*, and *SLC22A12*), and heritability
849 excluding index SNPs in all genome-wide significant loci for both the trans-ethnic and EA-
850 specific meta-analyses. Estimates were obtained by running 1,000,000 MCMC iterations (*burn*
851 *in* = 500,000) based on previously described settings[84]. The difference between the overall
852 heritability and the heritability excluding the index SNPs represents the heritability explained by
853 the significant loci in the present study.

854 **Trans-ethnic meta-regression**

855 Prior to conducting trans-ethnic meta-regression, we applied the same study-specific SNP filters
856 as those applied to the fixed effects trans-ethnic meta-analysis (imputation quality score >0.6
857 and MAC >10). An additional filter for MAF >0.0025 was also applied to reduce the influence of
858 very rare SNPs that passed the MAC filter in very large studies. Trans-ethnic meta-regression
859 was conducted using the MR-MEGA software package[85], which models ancestry-associated
860 heterogeneity in the allelic effect as a function of principal components (PCs) generated from a
861 matrix of mean pairwise allele frequency differences between studies. Three principal
862 components generated from a matrix of mean pairwise allele frequency differences between
863 studies were sufficient to separate the self-reported ancestry groups. Due to software
864 requirements, the minimum number of cohorts for each SNP had to be greater than the
865 number of PCs plus two, resulting in the exclusion of SNPs present in five or fewer cohorts. The
866 effect and p-value of each SNP on serum urate was reported, along with additional p-values
867 per-SNP for ancestry-associated heterogeneity ($p_{anc\text{-}het}$) and residual heterogeneity ($p_{res\text{-}het}$).
868 Index SNPs from the fixed effects meta-analysis with $p_{anc\text{-}het} < 2.7\times10^{-4}$ (0.05/183) in MR-MEGA
869 were considered to have significant ancestry-associated heterogeneity.

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

871 To evaluate the association of the trans-ethnic urate-associated index SNPs with gout, we
872 conducted a trans-ethnic meta-analysis of gout with the same study-specific filtering criteria as
873 for the urate trans-ethnic meta-analysis.

874 The association between a genetic urate risk score constructed from the 114
875 independent serum urate-associated SNPs identified among European individuals (see fine-
876 mapping section below) and gout was assessed in a large, independent sample from the UKBB
877 (Projects 19655 and 20272)[86]. We selected 334,880 unrelated individuals (pairwise kinship
878 coefficient<0.0313) of White British ancestry with sex chromosome euploidy and concordance
879 of phenotypic and genotypic sex, including 4,908 with gout identified by self-report at the
880 inclusion visit. Individuals with an ICD10 for gout (M10) in hospital admissions who did not self-
881 report gout were excluded from the analysis. A genetic risk score (GRS) was constructed as the
882 sum of the imputed dosage of the allele associated with higher urate levels ("risk alleles") over
883 all SNPs, multiplied by the genetic effect of the risk allele on serum urate levels. The GRS
884 distribution was divided into ten evenly-spaced categories, and individuals assigned to a
885 category based on their GRS. The category with the lowest GRS did not contain any gout cases
886 and so was combined with its adjacent category. Gout status was regressed on GRS category in
887 a logistic model, including age and sex as covariates, with the category containing the largest
888 number of individuals (genetically predicted mean urate levels 4.74-5.02 mg/dl higher
889 compared to individuals without any urate-increasing alleles) as the reference group.

890 The performance of the GRS for risk prediction of gout was trained on a set containing
891 90% of the sample and tested on the remaining 10%. Logistic regression was used to regress
892 gout on the GRS alone (genetic model), age and sex (demographic model) and GRS with age and
893 sex (combined model). Each of these models was then used to predict gout status in the test
894 set. Model performance was assessed by comparing predicted and true gout status using Area
895 Under the Curve (AUC) in a Receiver Operating Characteristic (ROC) curve. An optimum cutoff
896 of the ROC curve to report sensitivity and specificity of a combined GRS-based diagnostic test
897 was determined by the maximum of the Youden's index (sensitivity+specificity-1).
898

899 **Genetic correlation**

900 To assess the genetic correlation between serum urate and other traits in EA, we conducted
901 cross-trait LD score regression[87] using LD Hub[88] with the EA-specific urate meta-analysis results
902 as input. Genetic correlation estimates with 746 traits were obtained from LD Hub, excluding
903 two previous serum urate GWAS results. For presentation, the 212 significantly correlated traits
904 ($p<6.7x10^{-5}$=0.05/746) were grouped into 9 categories based on the trait names and labels and
905 presented in a circos plot.

906 To determine whether observed genetic correlations between serum urate and cardio-
907 metabolic traits are likely to represent causal relationships, we used the recently-developed
908 latent causal variable (LCV) method to estimate the genetic causality proportion (GCP) between
909 serum urate and another trait.[89] The GCP describes what proportion of the genetic component
910 of one trait also affects the other trait; a positive GCP value indicates that a large proportion of
911 the genetic component of urate also affects the other trait, and vice versa for a negative GCP
912 value. LCV produces posterior mean and standard deviation estimates of the GCP using mixed
913 fourth moments of the bivariate effect size distribution, based on GWAS summary statistics and
914 LD scores. When using summary statistics of cardio-metabolic traits generated from the UKBB,
915 we assumed non-overlapping populations, and overlapping populations otherwise. We selected
916 six unique continuous cardio-metabolic traits commonly examined in epidemiological studies
917 with high genetic correlation with serum urate ($|r_g|>0.35$). We additionally included gout as a
918 positive control and creatinine-based glomerular filtration rate. EA-specific GWAS summary
919 statistics were used as input to match the ancestry of the LD scores used with the method
920 (https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2).

921 **Functional enrichment**

922 To assess gene-set and tissue enrichment, we used the Data-Driven Expression Prioritized
923 Integration for Complex Traits analysis (DEPICT) version 1 release 194[90]. DEPICT performs gene
924 set enrichment analysis by testing whether genes in 14,461 reconstituted gene sets were
925 enriched for urate-associated SNPs. These reconstituted gene sets were generated based on
926 similarity analysis from gene expression of 77,840 samples, manually curated gene-sets,

927 molecular pathways from protein-protein interaction screening, and gene sets from mouse
928 gene knock-out studies. Additionally, gene expression levels of from 37,427 human microarray
929 samples are compiled in DEPICT and the tissue and cell types were mapped to 209 MeSH first
930 level terms including physiological systems, tissues and cells. Enrichment of tissues and cell
931 types was conducted by assessing whether urate-associated genes are highly expressed in the
932 tissue and cell types.

933 All variants with urate association p-values $<1 \times 10^{-5}$ from the trans-ethnic meta-analysis
934 results were used as input. Independent SNPs were identified using PLINK 1.9[74] clump
935 command within 500 kb flanking regions and $r^2>0.1$ in the 1000 Genomes phase1 version 3 data
936 excluding the MHC region (chr6:25–35 Mb). False discovery rates (FDRs) were computed using
937 500 repetitions, and p-values were computed using 5,000 permutations from 500 null GWAS
938 sets adjusting for gene length.

939 Affinity propagation clustering (APC)[91], implemented in the R package 'APCluster'[92], was
940 applied to all reconstituted gene sets with FDR-corrected enrichment P-value <0.01 to cluster
941 gene sets containing similar combinations of genes. Similarity between sets was assessed by the
942 z-scores of the top ten genes assigned to each gene set. The algorithm identifies a single
943 reconstituted gene set from each cluster as an exemplar (meta gene set) that best represents
944 the reconstituted gene sets within that cluster. A correlation matrix was calculated from the Z-
945 scores of the top ten genes within the meta gene set. Correlations >0.2 were visualized as
946 edges in a network in Cytoscape (http://cytoscape.org).

947 **Stratified LD score regression for functional enrichment**

948 Stratified LD score regression[87] estimates the SNP heritability of urate contributed by the SNPs
949 linked to histone marks in each cell type. The enrichment of a category is defined as the
950 proportion of SNP heritability in that cell type divided by the proportion of SNPs in the same cell
951 type. Here, we assessed urate heritability enrichment in 10 cell types via stratified LD score
952 regression with the EA-specific urate meta-analysis results as the input to match the ancestry of
953 the LD score estimates. The 10 cell types were collapsed from 220 cell type-specific annotations
954 for four histone marks: H3K4me1, H3K4me3, H3K9ac, and H3K27ac. Analyses were also carried

955 out using trans-ethnic meta-analysis summary statistics as input and results were similar (data
956 not shown).

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

958 Statistical fine-mapping to identify potentially causal variants was performed for the genome-
959 wide significant loci from the EA-specific meta-analysis, because fine-mapping based on
960 summary statistics relies on LD estimates from an ancestry-matched reference panel whose
961 sample size should scale with that of the GWAS (Methods)[34]. We estimated LD from genotypes
962 of 15,000 randomly selected UKBB participants (Application ID 2027, Dataset ID 8974). First, we
963 removed individuals who withdrew consent, those with mismatched sex between self-reported
964 and genetic, with genotype call rates <95%, outliers of variant heterozygosity or along the first
965 two principal components from a principal component analysis seeded with the HapMap phase
966 3 release 2 populations. We retained only one member of each pair of individuals with pair-
967 wise identity-by-descent statistic ≥0.1875. Altogether, 13,558 individuals with 16,969,363 SNPs
968 were selected as the LD reference panel for fine-mapping.

969 Statistical fine-mapping was conducted by first combining neighboring loci with
970 correlated index SNPs ($r^2 \geq 0.2$) in the EA meta-analysis into discrete regions. Next, we
971 performed stepwise model selection using GCTA (cojo-slct option) to identify independent
972 index SNPs in each region. For regions with more than one independent index SNPs, we
973 performed conditional analysis (GCTA cojo-cond option) to obtain conditional beta and
974 standard errors. Approximate Bayes factors (ABF) were calculated using the Wakefield's
975 formula[35], as implemented in the R package 'gtx' version 2.0.1
976 (https://github.com/tobyjohnson/gtx) using the conditional betas and standard errors for
977 regions with multiple independent SNPs and the betas and standard errors of the original EA
978 meta-analysis for regions with a single independent index SNP. The prior standard deviation
979 was calculated as 0.061 based on formula (8) of the original publication of the Wakefield's
980 formula (ref) and the 95% interval of the SNP effect sizes in the EA meta-analysis. The posterior
981 probability for a variant being the driver of the association signal was calculated as the ABF of
982 the variant divided by the sum of the ABF in the region. The 99% credible sets of a region is

983 derived by summing the posterior probabilities in descending order until the cumulative
984 posterior probability was >99%.

985 **Annotation of the variants in the credible sets**

986 We annotated SNPs in the credible sets for their exonic effect, Combined Annotation
987 Dependent Depletion (CADD) score, and mapping into DNaseI-hypersensitive sites (DHS) from
988 the Encyclopedia of DNA Elements (ENCODE) and Roadmap Epigenomics Consortium
989 projects[93,94]. The exonic effect and CADD score were obtained using SNiPA v3.2 (March 2017)[95].
990 SNiPA presented the CADD score as PHRED-like transformation of the C score, which was based
991 on CADD release v1.3 downloaded from http://cadd.gs.washington.edu/download. A CADD
992 score of 15 is used to distinguish potential pathogenic variants from background noise in clinical
993 genetics, and represents the median value of all non-synonymous variants in CADD v1.0[96,97].

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

995 Co-localization analysis of urate-associated loci with gene expression was conducted using EA-
996 specific urate meta-analysis results, *cis*-eQTL results from micro-dissected human glomerular
997 and tubulo-interstitial kidney portions from 187 individuals in the NEPTUNE study[98], as well as
998 from 44 tissues in the GTEx Project version 6p release[99]. For each locus, we identified all
999 transcripts and all tissue-transcript pairs with reported eQTLs within ±100 kb of each GWAS
1000 index SNP. The region for each co-localization test was defined as the eQTL *cis* window in the
1001 underlying studies[98,99]. We used the default parameters and prior definitions set in the
1002 'coloc.fast' function from the R package 'gtx' (https://github.com/tobyjohnson/gtx), which is an
1003 adapted implementation of Giambartolomei's colocalization method[24]. Evidence for co-
1004 localization was defined as H4 ≥0.8, which represents the posterior probability that the
1005 association with serum urate and gene expression is due to the same underlying variant. In
1006 addition, co-localization of urate-associated loci was also performed with gene expression
1007 quantified using RNA sequencing of the healthy tissue portion of 99 kidney cortex samples from
1008 the Cancer Genome Atlas (TCGA)[100]. First, all transcripts that shared eQTL variants with urate
1009 index SNPs within ±100 kb were extracted. Then the posterior probability of co-localization was

1010 calculated including eQTLs within the *cis*-window (±1Mb from the transcription start site) for
1011 each gene using the R coloc package[24] with default values for the three prior probabilities.

1012 **Trans-eQTL annotation**

1013 We performed trans-eQTL annotation for the 183 trans-ethnic index SNPs and their proxies (r^2
1014 >0.8 in both the 1000 Genomes phase 3 European and East Asian reference panels) in eQTL
1015 studies with >1,000 individuals: the Framingham Heart Study[101], Westra et al.[102], Zeller et al.[103],
1016 Fehrmann et al.[104], and the LIFE Heart[105] and LIFE-Adult[106] studies (details in Supplementary
1017 Methods). To improve stringency of results, we only report inter-chromosomal trans-eQTLs
1018 showing gene expression association p-values $<5x10^{-8}$ in at least two of the above mentioned
1019 independent samples. To characterize whether the identified trans-associated eQTL transcripts
1020 were enriched in any biological pathways, we conducted gene enrichment analysis using DOSE,
1021 an R/Bioconductor package for disease ontology semantic and enrichment analysis[107], as well as
1022 packages ReactomePA and clusterProfiler for enrichment analysis using the Human Disease
1023 Ontology database, GO, KEGG, and Reactome.[108-112] The background included all 19,327
1024 protein-coding genes from Ensembl/Havana reported in Ensembl release 91.

1025

1026 **Experimental study**

1027 ***Promoter Binding Site Predictions***

1028 For promoter binding site predictions, we used the JASPAR 2018 database[113,114]. The frequency
1029 matrices were downloaded for transcription factor binding sites of both vertebrate and human
1030 sequences (HNF1A: MA0046.1 and MA0046.2; HNF4A: MA0114.1 and MA0114.2). These
1031 matrices were then used to query the promoter region of *ABCG2* (-1285/+362, or base pairs
1032 upstream of the transcription start site / and downstream after transcription start site)[115] by
1033 means of the LASAGNA 2.0 transcription factor binding site search tool with default parameters
1034 and a p-value cutoff of 0.01[116].

1035 ***Site-Directed Mutagenesis***

1036 HNF1A and HNF4A clones were purchased from GeneCopoeia, (EX-A7792-M02 and EX-Z5283-
1037 M02 respectively) and were mutagenized using the QuikChange Lightning Site Directed
1038 Mutagenesis kit (Agilent Technologies, #210518) per manufacturer's instructions using PAGE
1039 purified primers.

1040 HNF1A-A98V-Forward: CCCTGAGGAGGCGGTCCACCAGAAAGCCG;

1041 HNF1A-A98V-Reverse: CGGCTTTCTGGTGGACCGCCTCCTCAGGG;

1042 HNF4A-T139I-Forward: GACCGGATCAGCATTCGAAGGTCAAGC;

1043 HNF4A-T139I-Reverse: GCTTGACCTTCGAATGCTGATCCGGTC.

1044 ***Luciferase Assay***

1045 HEK293T cells were seeded in white-walled 96-well plates coated with Poly-L-lysine at roughly
1046 12,500 cells per well. Cells were transfected 18 hours later with either the ABCG2 promoter (-
1047 1285/+362) upstream of a firefly luciferase in the pGL4.14 vector (a generous gift from Douglas
1048 D. Ross, University of Maryland School of Medicine), or the pGL4.14 vector (Promega, #E699A)
1049 without promoter construct, as well as GFP expressing vector used as an internal negative
1050 control (pEGFP-C1, Clontech)[117] using X-tremeGene™ 9 DNA Transfection Reagent (Roche
1051 Diagnostics, #6365787001). Transfection cocktails were prepared per manufacturer's
1052 specifications either with or without transcription factor using the following ratio: 0.6 μg
1053 promoter construct, 0.2, 0.1, or 0.05 μg transcription factor, and 0.05 μg GFP. When no
1054 transcription factor was used, pcDNA3.1 was substituted. Approximately 48 hours after
1055 transfection, cells were rinsed with 1x PBS, then lysed using Passive Lysis Buffer (Promega
1056 #E194A) for 15 minutes. During this incubation, GFP measurements were taken using a
1057 CLARIOstar microplate reader (BMG Labtech). Next, 30 μl of Luciferase Reagent (Promega,
1058 E297A&B) were added to each well, and the plate was incubated for an additional 20 minutes
1059 at room temperature. Finally, luciferase activity was measured using the CLARIOstar microplate
1060 reader taking the average over 6 seconds. To evaluate the significance of transactivation of the
1061 ABCG2 promoter we compared cells expressing transcription factors to those transfected with
1062 the empty vector (pcDAN3.1) and to evaluate TF dose responses or differences in TF variants all

1063 experimental conditions from one plate were compared using an Ordinary one-way ANOVA,
1064 accounting for multiple comparisons with a Tukey's multiple comparison test. Statistical
1065 analysis was performed using Prism 7 (GraphPad Software Inc, USA).

1066 ***Western Blots***

1067 Equal volumes of deoxycholate-RIPA buffer were added to wells containing desired lysates
1068 following the luciferase assay and plates were then incubated at 4°C overnight. Equal volumes
1069 of sample + 5x SDS loading dye + 10% β-merceptoethanol were then loaded into 10% Mini-
1070 PROTEAN® TGX Stain-Free™ Precast Gels (Bio-Rad, #4568033) and run per manufacturer's
1071 specifications. Gels were then cross-linked for 45 seconds and imaged to reveal total protein
1072 load, which was used as the loading control for each lane (representative images of these
1073 protein gels are found in S.upplementary figure 9). Gels were then transferred onto
1074 nitrocellulose membranes using the Trans-Blot® Turbo™ Transfer System (Bio-Rad), blocked for
1075 2 hours at room temperature in 5% milk in TBS-T, and incubated overnight at 4°C with primary
1076 antibody. Membranes were then washed 3 times with TBS-T, incubated at room temperature
1077 for 1 hour with Donkey anti-rabbit secondary antibody (Jackson ImmunoResearch, #111-035-
1078 144) diluted 1:5000 in 2.5% milk in TBS-T. Membranes were then washed again and developed
1079 using SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Thermo Scientific, #34577)
1080 and imaged on the ChemiDoc MP imaging system (Bio-Rad). All primary antibodies were diluted
1081 1:1000 in 2.5% milk in TBS-T. Antibodies used included HNF4α (Cell Signaling Technology,
1082 #3113) and HNF1α (Cell Signaling Technology, #89670). Antibodies were validated using lysates
1083 of overexpressing HEK293T cells transfected with either HNF construct, demonstrating bands at
1084 the appropriate sizes (Supplementary Figure 9).

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, DNase-I hypersensitivity site; CADD, Combined Annotation Dependent Depletion phred score; EA, European ancestry.

1089 **Acknowledgements**

1090 We thank Daniele Di Domizio (Eurac Research) for IT assistance, and Toby Johnson (GSK) for
1091 sharing his code and discussion on credible set fine-mapping and co-localization analysis. This
1092 research has been conducted using the UK Biobank Resource under Application Number 20272.
1093 Study-specific acknowledgements and funding sources are listed in the **Supplementary**
1094 **Information**.

1095

1096 **Disclaimer**

1097 The views expressed in this manuscript are those of the authors and do not necessarily
1098 represent the views of the National Heart, Lung, and Blood Institute, the National Institutes of
1099 Health, or the US Department of Health and Human Services.

1100

1101 **Data Availability**

1102 Genome-wide summary statistics for this study are made publicly available through dbGaP
1103 accession number phs000930.v7.p1.

1104

1105

1106 ## Author contributions

1107 **Manuscript writing group**: Adrienne Tin, Jonathan Marten, Victoria L. Halperin Kuhns, Yong Li, Matthias
1108 Wuttke, Holger Kirsten, Karsten B. Sieber, Chengxiang Qiu, Mathias Gorski, Markus Scholz, Adriana M.
1109 Hung, Alexander Teumer, Cristian Pattaro, Owen M. Woodward, Veronique Vitart, Anna Köttgen.
1110
1111 **Design of this study**: Adrienne Tin, Jonathan Marten, Matthias Wuttke, Mathias Gorski, Christian
1112 Fuchsberger, Alexander Teumer, Cristian Pattaro, Owen M. Woodward, Veronique Vitart, Anna Köttgen,
1113
1114 **Management of an individual contributing study:** Adam S. Butterworth, Adriana M. Hung, Adrienne Tin,
1115 Afshin Parsa, Aiko P.J. de Vries, Alan B. Zonderman, Alessandro De Grandi, Andres Metspalu, Andrew A.
1116 Hicks, Anke Tönjes, Anna Köttgen, Annette Peters, Antje Körner, Antonietta Robino, Archie Campbell,
1117 Belen Ponte, Bernhard K. Krämer, Bettina Jung, Brenda W.J.H. Penninx, Bruce M. Psaty, Caroline
1118 Hayward, Carsten A. Böger, Cassandra N. Spracklen, Christian Gieger, Christopher J. O'Donnell, Cornelia
1119 M. van Duijn, Cristian Pattaro, Daniela Toniolo, Daniele Cusi, Deborah Mascalzoni, Eric Boerwinkle, Erik
1120 Ingelsson, Florian Kronenberg, Gardar Sveinbjornsson, Georg Ehret, Gerard Waeber, Ginevra Biino,
1121 Girish N. Nadkarni, Grant W. Montgomery, Harold Snieder, Helena Schmidt, Igor Rudan, J. Michael
1122 Gaziano, James F. Wilson, James G. Wilson, Jaspal S. S. Kooner, Jeffrey OConnell, Joachim Thiery,
1123 Johanne Tremblay, John B. Whitfield, John C. Chambers, Josef Coresh, Kai-Uwe Eckardt, Karen L.
1124 Mohlke, Kari Stefansson, Kevin Ho, Koichi Matsuda, Konstantin Strauch, M. Arfan Ikram, Marcus E.
1125 Kleber, Marina Ciullo, Mario Pirastu, Markus Loeffler, Markus Scholz, Martin H. de Borst, Matthias
1126 Wuttke, Michael Stumvoll, Michele K. Evans, Michiaki Kubo, Mika Kähönen, Murielle Bochud, Myriam
1127 Rheinberger, Nicholas G. Martin, Olivier Devuyst, Olli T. Raitakari, Ozren Polasek, Paolo Gasparini, Peter
1128 P. Pramstaller, Peter Vollenweider, Pim van der Harst, Qiong Yang, Rainer Rettig, Reinhold Schmidt,
1129 RenÕ©e de Mutsert, Robert J. Carroll, Ron T. Gansevoort, Ruth J.F. Loos, Sarah A. Pendergrass, Sarah H.
1130 Wild, Stephan J.L. Bakker, Tamara B. Harris, Terho Lehtimäki, Thomas Perls, Ton J. Rabelink, Uwe Völker,
1131 Vilmantas Giedraitis, Vilmundur Gudnason, Weihua Zhang, Wieland Kiess, Winfried März, Wolfgang
1132 Koenig, Yong Li, Yuri Milaneschi
1133
1134 **Critical review of manuscript**: Adam S. Butterworth, Adriana M. Hung, Adrienne Tin, Afshin Parsa, Aiko
1135 P.J. de Vries, Alan B. Zonderman, Albert V. Smith, Alexander Teumer, André G. Uitterlinden, Anke
1136 Tönjes, Anna Köttgen, Annette Peters, Anselm Hoppmann, Antje Körner, Antonietta Robino, Anubha
1137 Mahajan, Audrey Y. Chu, Ayush Giri, Bernhard K. Krämer, Bettina Jung, Boting Ning, Bram Prins, Brenda
1138 W.J.H. Penninx, Brigitte Kühnel, Bruce M. Psaty, Caroline Hayward, Carsten A. Böger, Cassandra N.
1139 Spracklen, Chengxiang Qiu, Christa Meisinger, Christian Fuchsberger, Christian Gieger, Christopher J.
1140 O'Donnell, Cristian Pattaro, Daniel F. Gudbjartsson, Daniela Ruggiero, Deborah Mascalzoni, Dennis O.
1141 Mook-Kanamori, Erik Ingelsson, Erwin P. Bottinger, Eulalia Catamo, Florian Kronenberg, Gardar
1142 Sveinbjornsson, Ginevra Biino, Giorgia Girotto, Girish N. Nadkarni, Graciela Delgado, Grant W.
1143 Montgomery, Harold Snieder, Harry Campbell, , Helgi Jonsson, Hilma Holm, Igor Rudan, Ilja M. Nolte,
1144 Ingileif Jonsdottir, Iris M. Heid, James F. Wilson, James G. Wilson, Johanna Jakobsdottir, Johanne
1145 Tremblay, John B. Whitfield, Jonathan Marten, Josef Coresh, Kai-Uwe Eckardt, Karen L. Mohlke, Karlhans
1146 Endlich, Karsten B. Sieber, Katalin Susztak, Kenneth M. Rice, Kevin Ho, Kjell Nikus, Konstantin Strauch,
1147 Laura M. Raffield, Leo-Pekka Lyytikäinen, Leslie A. Lange, Luke J. O'Connor, Man Li, Marcus E. Kleber,
1148 Marina Ciullo, Markus Loeffler, Markus Scholz, Martin H. de Borst, Martina La Bianca, Martina Müller-
1149 Nurasyid, Mary L. Biggs, Mathias Gorski, Matthias Nauck, Matthias Wuttke, Melanie Waldenberger,
1150 Michael H. Preuss, Michele K. Evans, Mika Kähönen, Mike A. Nalls, Myriam Rheinberger, Nicholas G.
1151 Martin, Niek Verweij, Nina Hutri-Kähöne, Nisha Bansal, Olivier Devuyst, Olli T. Raitakari, Otis D. Wilson,

1152 Ozren Polasek, Patrick Sulem, Pavel Hamet, Peter K. Joshi, Pim van der Harst, Qiong Yang, Rainer Rettig,
1153 Ravchel M. Lewis, Raymond Noordam, RenÕ©e de Mutsert, Ruth J.F. Loos, Sahar Ghasemi, Sala Cinzia
1154 Felicita, Salman M. Tajuddin, Sanaz Sedaghat, Sarah A. Pendergrass, Sarah H. Wild, Scott D. Gordon,
1155 Shih-Jen Hwang, Shona M. Kerr, Stephan J.L. Bakker, Tamara B. Harris, Teresa Nutile, Terho Lehtimäki,
1156 Thibaud S. Boutin, Thomas Meitinger, Todd L. Edwards, Ton J. Rabelink, Unnur Thorsteinsdottir, Uwe
1157 Völker, Veronique Vitart, Wei Huang, Winfried März, Wolfgang Koenig, Yong Li, Zhi Yu.
1158

1159 **Statistical Methods and Analysis**: Albert V. Smith, Alexander Teumer, Anna Köttgen, Anselm
1160 Hoppmann, Anubha Mahajan, Audrey Y. Chu, Ayse Demirkan, Ayush Giri, Bettina Jung, Boting Ning,
1161 Bram Prins, Brigitte Kühnel, Carsten A. Böger, Cassandra N. Spracklen, Chengxiang Qiu, Chris H. L. Thio,
1162 Christian Fuchsberger, Cristian Pattaro, Damia Noce, Daniel F. Gudbjartsson, Edith Hofer, Erika Salvi,
1163 Federica Rizzi, Gardar Sveinbjornsson, Ginevra Biino, Graciela Delgado, Holger Kirsten, Ilja M. Nolte, Iris
1164 M. Heid, Johanna Jakobsdottir, Johanne Tremblay, Jonathan Marten, Jun Liu, Karsten B. Sieber, Katalin
1165 Susztak, Kathleen A. Ryan, Katrin Horn, Kenneth M. Rice, Laura M. Raffield, Leo-Pekka Lyytikäinen, Leslie
1166 A. Lange, Man Li, Marco Brumat, Marcus E. Kleber, Maria Pina Concas, Markus Scholz, Martin Gögele,
1167 Mary L. Biggs, Masahiro Kanai, Masato Akiyama, Massimiliano Cocca, Mathias Gorski, Matthias Nauck,
1168 Matthias Wuttke, Michael H. Preuss, Mike A. Nalls, Myriam Rheinberger, Navya Shilpa Josyula, Nicola
1169 Pirastu, Niek Verweij, Nina Mononen, Pashupati P. Mishra, Pavel Hamet, Peter J. van der Most, Peter K.
1170 Joshi, Pim van der Harst, Qiong Yang, Raymond Noordam, Rico Rueedi, Robert J. Carroll, Sahar Ghasemi,
1171 Salman M. Tajuddin, Sanaz Sedaghat, Sarah A. Pendergrass, Shih-Jen Hwang, Tanguy Corre, Teresa
1172 Nutile, Thibaud S. Boutin, Todd L. Edwards, Toomas Haller, Veronique Vitart, Weihua Zhang, Winfried
1173 März, Yasaman Saba, Yizhe Xu, Yoichiro Kamatani, Yong Li, Yukinori Okada
1174

1175 **Subject Recruitment:** Aiko P.J. de Vries, Alan B. Zonderman, Andrej Teren, Andres Metspalu, Anke
1176 Tönjes, Anna Köttgen, Archie Campbell, Belen Ponte, Bettina Jung, Blair H. Smith, Brenda W.J.H.
1177 Penninx, Carsten A. Böger, Christa Meisinger, Cristian Pattaro, Daniela Ruggiero, Daniele Cusi, David J.
1178 Porteous, Erwin P. Bottinger, Florian Kronenberg, Gerard Waeber, Harry Campbell, Helgi Jonsson, Igor
1179 Rudan, Isleifur Olafsson, James F. Wilson, James G. Wilson, Jaspal S. S. Kooner, Johan Ärnlöv, Johanne
1180 Tremblay, John B. Whitfield, John C. Chambers, Katalin Dittrich, Kjell Nikus, Koichi Matsuda, Marina
1181 Ciullo, Michele K. Evans, Michiaki Kubo, Mika Kähönen, Myriam Rheinberger, Nicholas G. Martin, Nina
1182 Hutri-Kähöne, Olli T. Raitakari, Ozren Polasek, Patrick Sulem, Peter Vollenweider, Reinhold Schmidt,
1183 Renée de Mutsert, Ron T. Gansevoort, Saima Afaq, Sandosh Padmanabhan, Sarah A. Pendergrass, Sarah
1184 H. Wild, Simona Vaccargiu, Tanja Poulain, Terho Lehtimäki, Ton J. Rabelink, Vilmundur Gudnason, Wei
1185 Huang, Winfried März
1186

1187 **Bioinformatics**: Albert V. Smith, Anna Köttgen, Anselm Hoppmann, Audrey Y. Chu, Ayush Giri, Benjamin
1188 Lehne, Bram Prins, Carsten A. Böger, Cassandra N. Spracklen, Chengxiang Qiu, Christian M. Shaffer,
1189 Daniela Baptista, Dennis O. Mook-Kanamori, Edith Hofer, Eric Campana, Erika Salvi, Federica Rizzi, Georg
1190 Ehret, Giorgio Pistis, Holger Kirsten, Iris M. Heid, James F. Wilson, Johanna Jakobsdottir, Johanne
1191 Tremblay, Jonathan Marten, Karen L. Mohlke, Karsten B. Sieber, Katalin Susztak, Katrin Horn, Leo-Pekka
1192 Lyytikäinen, Man Li, Marcus E. Kleber, Maria Pina Concas, Markus Scholz, Massimiliano Cocca, Mathias
1193 Gorski, Matthias Wuttke, Michael H. Preuss, Navya Shilpa Josyula, Nicola Pirastu, Pashupati P. Mishra,
1194 Pavel Hamet, Peter J. van der Most, Raymond Noordam, Reedik Magi, Rico Rueedi, Robert J. Carroll,
1195 Sahar Ghasemi, Sanaz Sedaghat, Sarah A. Pendergrass, Scott D. Gordon, Sven Bergmann, Tanguy Corre,
1196 Teresa Nutile, Weihua Zhang, Winfried März, Yasaman Saba, Yizhe Xu, Yong Li, Yuri Milaneschi, Zhi Yu.
1197

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, 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, Luke J. O'Connor, 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 F. Wilson, 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, Raychel Lewis, Owen M. Woodward.

Competing interests

Dennis O Mook-Kanamori works as a part-time clinical research consultant for Metabolon, Inc. Brenda Penninx has received research funding (unrelated to the work reported here) from Jansen Research and Boehringer Ingelheim. 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. 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.

1242 **Figure Legends**

1243 **Figure 1: Trans-ethnic GWAS meta-analysis identifies 183 loci associated with serum urate**

1244 Outer ring: Dot size represents the genetic effect size of the index SNP at each labeled locus on
1245 serum urate. Blue band: $-\log_{10}(p)$ for association with serum urate, by chromosomal position
1246 (GRCh37 (hg19) reference build). Red line indicates genome-wide significance ($p=5\times10^{-8}$). Blue
1247 gene labels indicate novel loci, gray labels loci reported in previous GWAS of serum urate.
1248 Green band: $-\log_{10}(p)$ for association with gout, by chromosomal position. Red line indicates
1249 genome-wide significance ($p=5\times10^{-8}$). Inner band: Dots represent index SNPs with significant
1250 heterogeneity and are color-coded according to its source: green for ancestry-related
1251 heterogeneity ($p_{anc\text{-}het} < 2.7\times10^{-4}$ [0.05/183]), red for residual heterogeneity ($p_{res\text{-}het} < 2.7\times10^{-4}$),
1252 and yellow for both ($p_{anc\text{-}het}$ and $p_{res\text{-}het} < 2.7\times10^{-4}$). Loci are labeled with the gene closest to the
1253 index SNP.

1254

1255 **Figure 2: A genetic risk score (GRS) for serum urate improves gout risk prediction. (A)**
1256 Histogram of the urate GRS among 334,880 European ancestry participants of the UK Biobank.
1257 The Y axes show the number of individuals (left) and the prevalence of gout (right), the x-axis
1258 shows categories of the urate GRS. The units on the x-axis represent genetically-predicted
1259 serum urate levels (mg/dl) compared to individuals without any urate-increasing alleles; **(B)** Y
1260 axis displays the age- and sex-adjusted odds ratio of gout by GRS category (X axis), comparing
1261 each other category to the most prevalent one; **(C)** Comparison of the receiver operating
1262 characteristic (ROC) curves of different prediction models of gout: genetic (GRS only; red),
1263 demographic (age + sex; green), and combined (GRS + age + sex; blue). Y-axis: sensitivity, x-axis:
1264 specificity. At the optimal cut points determined by the maximum of the Youden's index, the
1265 sensitivity of the combined model was 84% and specificity was 68%.

1266

1267 **Figure 3: Serum urate shows widespread genetic correlations with cardio-metabolic risk**
1268 **factors and diseases.**

1269 The Circos plot shows significant genome-wide genetic correlations between serum urate and
1270 214 complex traits or diseases ($p<6.6x10^{-5}$), with bar height proportional to the genetic
1271 correlation coefficient (r_g) estimate for each trait and coloring according to its direction (dark
1272 blue, $r_g>0$; light blue, $r_g<0$). Traits and diseases are labeled on the outside of the plot and
1273 grouped into nine different categories. Each category is color-coded (inner ring, inset). The
1274 greatest genetic correlation was observed with gout ($r_g=0.92$, $p=3.3x10^{-70}$). Genetic correlations
1275 with multiple cardio-metabolic risk factors and diseases reflect their known directions from
1276 observational studies.

1277

1278 **Figure 4: Genes expressed in urate-associated loci are enriched in kidney tissue and**
1279 **pathways. (A)** Grouped physiological systems (x-axis) that were tested individually for
1280 enrichment of expression of genes in urate-associated loci are shown as a bar plot, with the –
1281 $\log_{10}$(*P*-value) on the y-axis. Significantly enriched systems are labeled and highlighted in blue
1282 (false discovery rate [FDR] <0.01). **(B)** Correlated (r > 0.2) meta-gene sets that were strongly
1283 enriched for genes mapping into urate-associated loci (FDR <0.01). Thickness of the edges
1284 represents the magnitude of the correlation coefficient, node size, color and intensity represent
1285 the number of clustered gene sets, gene set origin, and enrichment p-value, respectively.

1286

1287 **Figure 5: Prioritization of p.Thr139Ile at *HNF4A* and functional study of *HNF4A* regulation of**
1288 ***ABCG2* transcription.**

1289 **(A)** Graph shows credible set size (x-axis) against the posterior probability of association (PPA;
1290 y-axis) for each of 1,453 SNPs with PPA >1% in 114 99% credible sets. Triangles mark missense
1291 SNPs, with size proportional to their Combined Annotation Dependent Depletion (CADD) score.
1292 Blue triangles indicate missense variants mapping into small (≤5 SNPs) credible sets or with high
1293 PPA (≥50%). **(B)** Predicted HNF1A or HNF4A binding sites in the promoter region of *ABCG2*, the
1294 consensus affinity sequence, and the p-value of likely matches. **(C)** Relative luciferase activity
1295 and transactivation of *ABCG2* promoter in cells transfected with variable amount of HNF1A or

1296 HNF4A constructs. ± SEM, n=3 independent experiments, p values calculated with ordinary one-
1297 way ANOVA with Tukey's multiple comparison test. **(D)** Position of p.Thr139Ile (T139I) in DNA
1298 binding domain / hinge region within HNF4A homodimer structure (PDB 4IQR). **(E)** Relative
1299 luciferase activity and transactivation of *ABCG2* promoter in cells transfected with variable
1300 amount of constructs (ng's of transfected DNA) of wild-type HNF4A (threonine) or isoleucine at
1301 position 139. ± SEM, n=3 independent experiments, p values calculated with ordinary one-way
1302 ANOVA with Tukey's multiple comparison test.

1303

1304 **Figure 6: Co-localization of urate-association signals with gene expression in *cis* in kidney**
1305 **tissues**

1306 Serum urate association signals identified among European ancestry individuals were tested for
1307 co-localization with all eQTLs where the eQTL *cis*-window overlapped (±100 kb) the index SNP.
1308 Genes with ≥1 positive co-localization (posterior probability of one common causal variant, H4,
1309 ≥0.80) in a kidney tissue are illustrated with the respective index SNP and transcript (y-axis). Co-
1310 localizations across all tissues (x-axis) are illustrated as dots, where the size of the dots indicates
1311 the posterior probability of the co-localization. Negative co-localizations (posterior probability
1312 of H4 <0.80) are marked in gray, while the positive co-localizations are color-coded based on
1313 the predicted change in expression relative to the allele associated with higher serum urate.

1314

1315 **Table 1: Genes implicated as causal via identification of missense variants with high**
1316 **probability of driving the urate association signal.** Genes are included if they contain a
1317 missense variant with posterior probability of association of >50% or mapping into a small
1318 credible set (≤5 variants).

1319

1320 **References**

1321 1. Kuo, C.F., Grainge, M.J., Zhang, W. & Doherty, M. Global epidemiology of gout:
1322 prevalence, incidence and risk factors. *Nat Rev Rheumatol* **11**, 649-62 (2015).
1323 2. Li, X. *et al.* Serum uric acid levels and multiple health outcomes: umbrella review of
1324 evidence from observational studies, randomised controlled trials, and Mendelian
1325 randomisation studies. *BMJ* **357**, j2376 (2017).
1326 3. Jinno, S., Hasegawa, K., Neogi, T., Goto, T. & Dubreuil, M. Trends in Emergency
1327 Department Visits and Charges for Gout in the United States between 2006 and 2012. *J*
1328 *Rheumatol* **43**, 1589-92 (2016).
1329 4. Kuo, C.F., Grainge, M.J., Mallen, C., Zhang, W. & Doherty, M. Rising burden of gout in
1330 the UK but continuing suboptimal management: a nationwide population study. *Ann*
1331 *Rheum Dis* **74**, 661-7 (2015).
1332 5. Mikuls, T.R., Farrar, J.T., Bilker, W.B., Fernandes, S. & Saag, K.G. Suboptimal physician
1333 adherence to quality indicators for the management of gout and asymptomatic
1334 hyperuricaemia: results from the UK General Practice Research Database (GPRD).
1335 *Rheumatology (Oxford)* **44**, 1038-42 (2005).
1336 6. Yang, Q. *et al.* Genome-wide search for genes affecting serum uric acid levels: the
1337 Framingham Heart Study. *Metabolism* **54**, 1435-41 (2005).
1338 7. Vitart, V. *et al.* SLC2A9 is a newly identified urate transporter influencing serum urate
1339 concentration, urate excretion and gout. *Nat Genet* **40**, 437-42 (2008).
1340 8. Pilia, G. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians.
1341 *PLoS Genet* **2**, e132 (2006).
1342 9. Wang, W. *et al.* Heritability and Genome-Wide Association Analyses of Serum Uric Acid
1343 in Middle and Old-Aged Chinese Twins. *Front Endocrinol (Lausanne)* **9**, 75 (2018).
1344 10. MacCluer, J.W. *et al.* Heritability of measures of kidney disease among Zuni Indians: the
1345 Zuni Kidney Project. *Am J Kidney Dis* **56**, 289-302 (2010).
1346 11. Rule, A.D. *et al.* Genome-wide linkage analysis for uric acid in families enriched for
1347 hypertension. *Nephrol Dial Transplant* **24**, 2414-20 (2009).
1348 12. Enomoto, A. *et al.* Molecular identification of a renal urate anion exchanger that
1349 regulates blood urate levels. *Nature* **417**, 447-52 (2002).
1350 13. Li, S. *et al.* The GLUT9 gene is associated with serum uric acid levels in Sardinia and
1351 Chianti cohorts. *PLoS Genet* **3**, e194 (2007).
1352 14. Doring, A. *et al.* SLC2A9 influences uric acid concentrations with pronounced sex-
1353 specific effects. *Nat Genet* **40**, 430-6 (2008).
1354 15. Dehghan, A. *et al.* Association of three genetic loci with uric acid concentration and risk
1355 of gout: a genome-wide association study. *Lancet* **372**, 1953-61 (2008).
1356 16. Kolz, M. *et al.* Meta-analysis of 28,141 individuals identifies common variants within five
1357 new loci that influence uric acid concentrations. *PLoS Genet* **5**, e1000504 (2009).
1358 17. Yang, Q. *et al.* Multiple genetic loci influence serum urate levels and their relationship
1359 with gout and cardiovascular disease risk factors. *Circ Cardiovasc Genet* **3**, 523-30
1360 (2010).
1361 18. Tin, A. *et al.* Genome-wide association study for serum urate concentrations and gout
1362 among African Americans identifies genomic risk loci and a novel URAT1 loss-of-
1363 function allele. *Hum Mol Genet* **20**, 4056-68 (2011).
1364 19. Woodward, O.M. *et al.* Identification of a urate transporter, ABCG2, with a common
1365 functional polymorphism causing gout. *Proc Natl Acad Sci U S A* **106**, 10338-42 (2009).
1366 20. Major, T.J., Dalbeth, N., Stahl, E.A. & Merriman, T.R. An update on the genetics of
1367 hyperuricaemia and gout. *Nat Rev Rheumatol* **14**, 341-353 (2018).

21. Kottgen, A. *et al.* Genome-wide association analyses identify 18 new loci associated with serum urate concentrations. *Nat Genet* **45**, 145-54 (2013).
22. Kanai, M. *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nat Genet* **50**, 390-400 (2018).
23. Schaid, D.J., Chen, W. & Larson, N.B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* **19**, 491-504 (2018).
24. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* **10**, e1004383 (2014).
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. Merriman, T.R. Population heterogeneity in the genetic control of serum urate. *Semin Nephrol* **31**, 420-5 (2011).
28. Roddy, E. & Choi, H.K. Epidemiology of gout. *Rheum Dis Clin North Am* **40**, 155-75 (2014).
29. 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).
30. Sautin, Y.Y. & Johnson, R.J. Uric acid: the oxidant-antioxidant paradox. *Nucleosides Nucleotides Nucleic Acids* **27**, 608-19 (2008).
31. Long, W. *et al.* Identification of Key Residues for Urate Specific Transport in Human Glucose Transporter 9 (hSLC2A9). *Sci Rep* **7**, 41167 (2017).
32. Feig, D.I., Kang, D.H. & Johnson, R.J. Uric acid and cardiovascular risk. *N Engl J Med* **359**, 1811-21 (2008).
33. Lyngdoh, T. *et al.* Serum uric acid and adiposity: deciphering causality using a bidirectional Mendelian randomization approach. *PLoS One* **7**, e39321 (2012).
34. Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am J Hum Genet* **101**, 539-551 (2017).
35. Wakefield, J. A Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am J Hum Genet* **81**, 208-27 (2007).
36. Pao, S.S., Paulsen, I.T. & Saier, M.H., Jr. Major facilitator superfamily. *Microbiol Mol Biol Rev* **62**, 1-34 (1998).
37. Asano, T. *et al.* The role of N-glycosylation of GLUT1 for glucose transport activity. *J Biol Chem* **266**, 24632-6 (1991).
38. Boyle, E.A., Li, Y.I. & Pritchard, J.K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186 (2017).
39. Prestin, K. *et al.* Regulation of PDZ domain-containing 1 (PDZK1) expression by hepatocyte nuclear factor-1alpha (HNF1alpha) in human kidney. *Am J Physiol Renal Physiol* **313**, F973-F983 (2017).
40. Maher, J.M. *et al.* Alterations in transporter expression in liver, kidney, and duodenum after targeted disruption of the transcription factor HNF1alpha. *Biochem Pharmacol* **72**, 512-22 (2006).
41. Sulem, P. *et al.* Identification of low-frequency variants associated with gout and serum uric acid levels. *Nat Genet* **43**, 1127-30 (2011).
42. 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).

1418 43. Lin, C.I. *et al.* Modulation of the neuronal glutamate transporter EAAC1 by the interacting
1419 protein GTRAP3-18. *Nature* **410**, 84-8 (2001).
1420 44. Butchbach, M.E., Lai, L. & Lin, C.L. Molecular cloning, gene structure, expression profile
1421 and functional characterization of the mouse glutamate transporter (EAAT3) interacting
1422 protein GTRAP3-18. *Gene* **292**, 81-90 (2002).
1423 45. Bailey, C.G. *et al.* Loss-of-function mutations in the glutamate transporter SLC1A1 cause
1424 human dicarboxylic aminoaciduria. *J Clin Invest* **121**, 446-53 (2011).
1425 46. Merkulova, M. *et al.* Targeted deletion of the Ncoa7 gene results in incomplete distal
1426 renal tubular acidosis in mice. *Am J Physiol Renal Physiol* **315**, F173-F185 (2018).
1427 47. Kirby, A. *et al.* Mutations causing medullary cystic kidney disease type 1 lie in a large
1428 VNTR in MUC1 missed by massively parallel sequencing. *Nat Genet* **45**, 299-303
1429 (2013).
1430 48. Kraus, M.R. *et al.* Two mutations in human BICC1 resulting in Wnt pathway hyperactivity
1431 associated with cystic renal dysplasia. *Hum Mutat* **33**, 86-90 (2012).
1432 49. Hart, T.C. *et al.* Mutations of the UMOD gene are responsible for medullary cystic kidney
1433 disease 2 and familial juvenile hyperuricaemic nephropathy. *J Med Genet* **39**, 882-92
1434 (2002).
1435 50. Sodini, S.M., Kemper, K.E., Wray, N.R. & Trzaskowski, M. Comparison of Genotypic and
1436 Phenotypic Correlations: Cheverud's Conjecture in Humans. *Genetics* **209**, 941-948
1437 (2018).
1438 51. Lindgren, D. *et al.* Cell-Type-Specific Gene Programs of the Normal Human Nephron
1439 Define Kidney Cancer Subtypes. *Cell Rep* **20**, 1476-1489 (2017).
1440 52. Prestin, K. *et al.* Transcriptional regulation of urate transportosome member SLC2A9 by
1441 nuclear receptor HNF4alpha. *Am J Physiol Renal Physiol* **307**, F1041-51 (2014).
1442 53. Ketharnathan, S. *et al.* A non-coding genetic variant maximally associated with serum
1443 urate levels is functionally linked to HNF4A-dependent PDZK1 expression. *Hum Mol*
1444 *Genet* **27**, 3964-3973 (2018).
1445 54. Marable, S.S., Chung, E., Adam, M., Potter, S.S. & Park, J.S. Hnf4a deletion in the
1446 mouse kidney phenocopies Fanconi renotubular syndrome. *JCI Insight* **3**(2018).
1447 55. Matsuo, H. *et al.* ABCG2 dysfunction causes hyperuricemia due to both renal urate
1448 underexcretion and renal urate overload. *Sci Rep* **4**, 3755 (2014).
1449 56. Daigo, K. *et al.* Proteomic analysis of native hepatocyte nuclear factor-4alpha
1450 (HNF4alpha) isoforms, phosphorylation status, and interactive cofactors. *J Biol Chem*
1451 **286**, 674-86 (2011).
1452 57. Chandra, V. *et al.* Multidomain integration in the structure of the HNF-4alpha nuclear
1453 receptor complex. *Nature* **495**, 394-8 (2013).
1454 58. Zhu, Q. *et al.* T130I mutation in HNF-4alpha gene is a loss-of-function mutation in
1455 hepatocytes and is associated with late-onset Type 2 diabetes mellitus in Japanese
1456 subjects. *Diabetologia* **46**, 567-73 (2003).
1457 59. Fuchsberger, C., Taliun, D., Pramstaller, P.P. & Pattaro, C. GWAtoolbox: an R package
1458 for fast quality control and handling of genome-wide association studies meta-analysis
1459 data. *Bioinformatics* **28**, 444-5 (2012).
1460 60. Devlin, B., Roeder, K. & Wasserman, L. Genomic control, a new approach to genetic-
1461 based association studies. *Theor Popul Biol* **60**, 155-66 (2001).
1462 61. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics*
1463 *Hum Genet* **10**, 387-406 (2009).
1464 62. Delaneau, O., Marchini, J. & Zagury, J.F. A linear complexity phasing method for
1465 thousands of genomes. *Nat Methods* **9**, 179-81 (2011).
1466 63. Loh, P.R., Palamara, P.F. & Price, A.L. Fast and accurate long-range phasing in a UK
1467 Biobank cohort. *Nat Genet* **48**, 811-6 (2016).

1468 64. Browning, B.L. & Browning, S.R. A fast, powerful method for detecting identity by
1469 descent. *Am J Hum Genet* **88**, 173-82 (2011).
1470 65. the Haplotype Reference, C. A reference panel of 64,976 haplotypes for genotype
1471 imputation. *Nature Genetics* **48**, 1279 (2016).
1472 66. Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human
1473 genomes. *Nature* **491**, 56-65 (2012).
1474 67. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation
1475 method for the next generation of genome-wide association studies. *PLoS Genet* **5**,
1476 e1000529 (2009).
1477 68. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**,
1478 1284-1287 (2016).
1479 69. Durbin, R. Efficient haplotype matching and storage using the positional Burrows-
1480 Wheeler transform (PBWT). *Bioinformatics* **30**, 1266-72 (2014).
1481 70. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat*
1482 *Genet* **48**, 1279-83 (2016).
1483 71. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat*
1484 *Rev Genet* **11**, 499-511 (2010).
1485 72. Haller, T., Kals, M., Esko, T., Magi, R. & Fischer, K. RegScan: a GWAS tool for quick
1486 estimation of allele effects on continuous traits and their combinations. *Brief Bioinform*
1487 **16**, 39-44 (2015).
1488 73. Zhan, X., Hu, Y., Li, B., Abecasis, G.R. & Liu, D.J. RVTESTS: an efficient and
1489 comprehensive tool for rare variant association analysis using sequence data.
1490 *Bioinformatics* **32**, 1423-6 (2016).
1491 74. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer
1492 datasets. *Gigascience* **4**, 7 (2015).
1493 75. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-
1494 wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).
1495 76. Chen, M.H. & Yang, Q. GWAF: an R package for genome-wide association analyses
1496 with family data. *Bioinformatics* **26**, 580-1 (2010).
1497 77. Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association
1498 studies. *Nat Genet* **44**, 821-4 (2012).
1499 78. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and
1500 genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* **34**,
1501 816-34 (2010).
1502 79. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of
1503 genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
1504 80. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat Med*
1505 **21**, 1539-58 (2002).
1506 81. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from
1507 polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
1508 82. Hadfield, J. MCMC methods for multi-response generalized linear mixed models: the
1509 MCMC glmm R Package. *J Stat Softw* **33**, 1-22 (2010).
1510 83. Pattaro, C. *et al.* The Cooperative Health Research in South Tyrol (CHRIS) study:
1511 rationale, objectives, and preliminary results. *J Transl Med* **13**, 348 (2015).
1512 84. Hadfield, J.D. MCMC Methods for Multi-Response Generalized Linear Mixed Models:
1513 The MCMCglmm R Package. *Journal of Statistical Software; Vol 1, Issue 2 (2010)*
1514 (2010).
1515 85. Magi, R. *et al.* Trans-ethnic meta-regression of genome-wide association studies
1516 accounting for ancestry increases power for discovery and improves fine-mapping
1517 resolution. *Hum Mol Genet* **26**, 3639-3650 (2017).

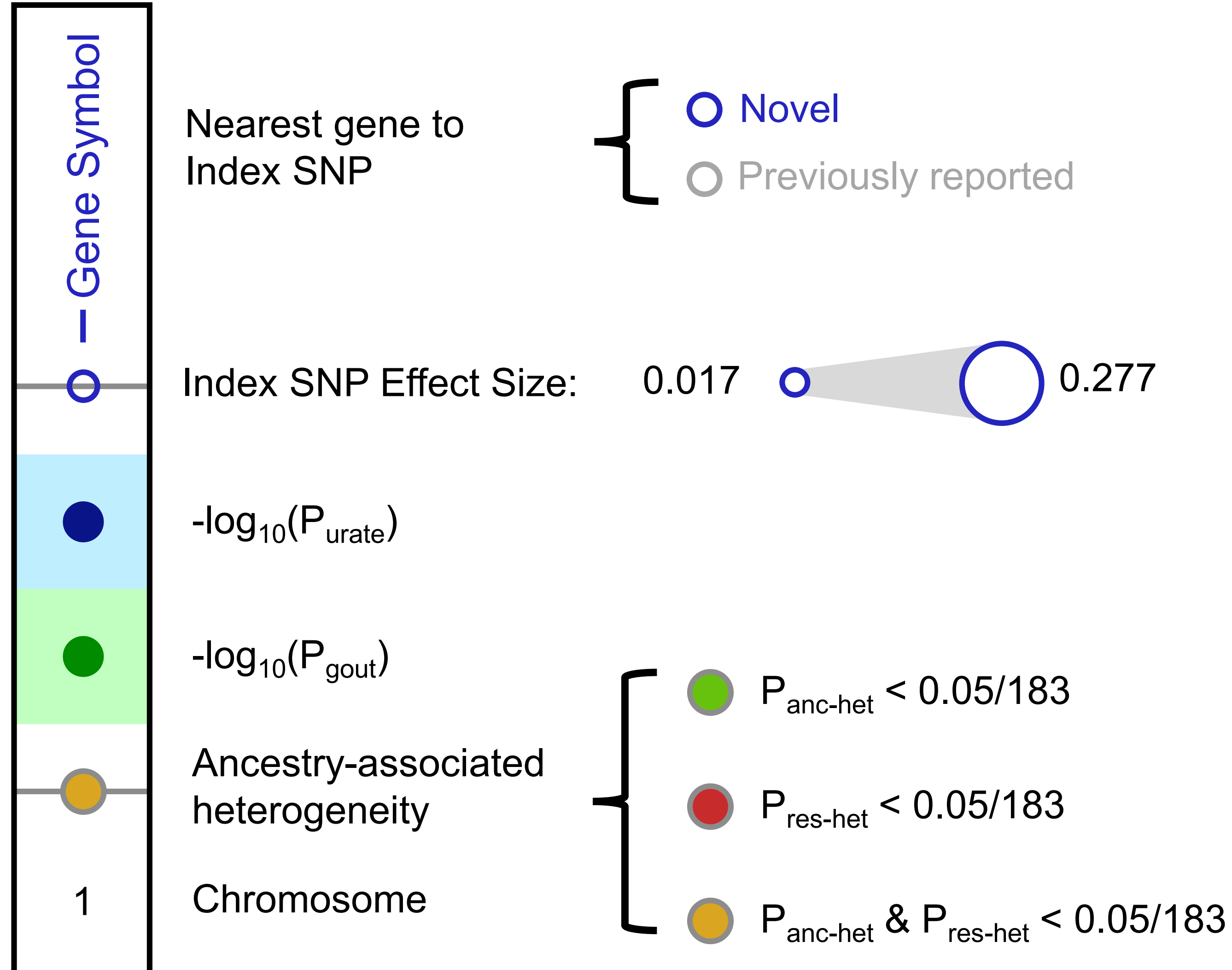
1518 86. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a
1519 wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
1520 87. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and
1521 traits. *Nat Genet* **47**, 1236-41 (2015).
1522 88. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score
1523 regression that maximizes the potential of summary level GWAS data for SNP
1524 heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279 (2017).
1525 89. O'Connor, L.J. & Price, A.L. Distinguishing genetic correlation from causation across 52
1526 diseases and complex traits. *Nat Genet* **50**, 1728-1734 (2018).
1527 90. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using
1528 predicted gene functions. *Nat Commun* **6**, 5890 (2015).
1529 91. Frey, B.J. & Dueck, D. Clustering by passing messages between data points. *Science*
1530 **315**, 972-6 (2007).
1531 92. Bodenhofer, U., Kothmeier, A. & Hochreiter, S. APCluster: an R package for affinity
1532 propagation clustering. *Bioinformatics* **27**, 2463-4 (2011).
1533 93. Sheffield, N.C. *et al.* Patterns of regulatory activity across diverse human cell types
1534 predict tissue identity, transcription factor binding, and long-range interactions. *Genome*
1535 *Res* **23**, 777-88 (2013).
1536 94. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**,
1537 317-30 (2015).
1538 95. Arnold, M., Raffler, J., Pfeufer, A., Suhre, K. & Kastenmuller, G. SNiPA: an interactive,
1539 genetic variant-centered annotation browser. *Bioinformatics* **31**, 1334-6 (2015).
1540 96. Dong, C. *et al.* Comparison and integration of deleteriousness prediction methods for
1541 nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet* **24**, 2125-
1542 37 (2015).
1543 97. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human
1544 genetic variants. *Nat Genet* **46**, 310-5 (2014).
1545 98. Gillies, C.E. *et al.* An eQTL Landscape of Kidney Tissue in Human Nephrotic Syndrome.
1546 *Am J Hum Genet* **103**, 232-244 (2018).
1547 99. GTEx. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-5 (2013).
1548 100. Ko, Y.A. *et al.* Genetic-Variation-Driven Gene-Expression Changes Highlight Genes with
1549 Important Functions for Kidney Disease. *Am J Hum Genet* **100**, 940-953 (2017).
1550 101. Joehanes, R. *et al.* Integrated genome-wide analysis of expression quantitative trait loci
1551 aids interpretation of genomic association studies. *Genome Biol* **18**, 16 (2017).
1552 102. Westra, H.J. *et al.* Systematic identification of trans eQTLs as putative drivers of known
1553 disease associations. *Nat Genet* **45**, 1238-1243 (2013).
1554 103. Zeller, T. *et al.* Genetics and beyond--the transcriptome of human monocytes and
1555 disease susceptibility. *PLoS One* **5**, e10693 (2010).
1556 104. Fehrmann, R.S. *et al.* Trans-eQTLs reveal that independent genetic variants associated
1557 with a complex phenotype converge on intermediate genes, with a major role for the
1558 HLA. *PLoS Genet* **7**, e1002197 (2011).
1559 105. Ichida, K. *et al.* Decreased extra-renal urate excretion is a common cause of
1560 hyperuricemia. *Nat Commun* **3**, 764 (2012).
1561 106. Qu, L.H., Jiang, H. & Chen, J.H. Effect of uric acid-lowering therapy on blood pressure:
1562 systematic review and meta-analysis. *Ann Med* **49**, 142-156 (2017).
1563 107. Yu, G., Wang, L.G., Yan, G.R. & He, Q.Y. DOSE: an R/Bioconductor package for
1564 disease ontology semantic and enrichment analysis. *Bioinformatics* **31**, 608-9 (2015).
1565 108. Schriml, L.M. *et al.* Human Disease Ontology 2018 update: classification, content and
1566 workflow expansion. *Nucleic Acids Res* **47**, D955-D962 (2019).

1567 109. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res* **45**,
1568 D331-D338 (2017).
1569 110. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene
1570 Ontology Consortium. *Nat Genet* **25**, 25-9 (2000).
1571 111. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new
1572 perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353-
1573 D361 (2017).
1574 112. Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Res* **46**,
1575 D649-D655 (2018).
1576 113. Khan, A. *et al.* JASPAR 2018: update of the open-access database of transcription
1577 factor binding profiles and its web framework. *Nucleic Acids Res* **46**, D1284 (2018).
1578 114. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W. & Lenhard, B. JASPAR: an
1579 open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids*
1580 *Res* **32**, D91-4 (2004).
1581 115. Xie, Y. *et al.* Functional cyclic AMP response element in the breast cancer resistance
1582 protein (BCRP/ABCG2) promoter modulates epidermal growth factor receptor pathway-
1583 or androgen withdrawal-mediated BCRP/ABCG2 transcription in human cancer cells.
1584 *Biochim Biophys Acta* **1849**, 317-27 (2015).
1585 116. Lee, C. & Huang, C.H. LASAGNA-Search 2.0: integrated transcription factor binding site
1586 search and visualization in a browser. *Bioinformatics* **30**, 1923-5 (2014).
1587 117. Vesuna, F., Winnard, P., Jr. & Raman, V. Enhanced green fluorescent protein as an
1588 alternative control reporter to Renilla luciferase. *Anal Biochem* **342**, 345-7 (2005).

1589

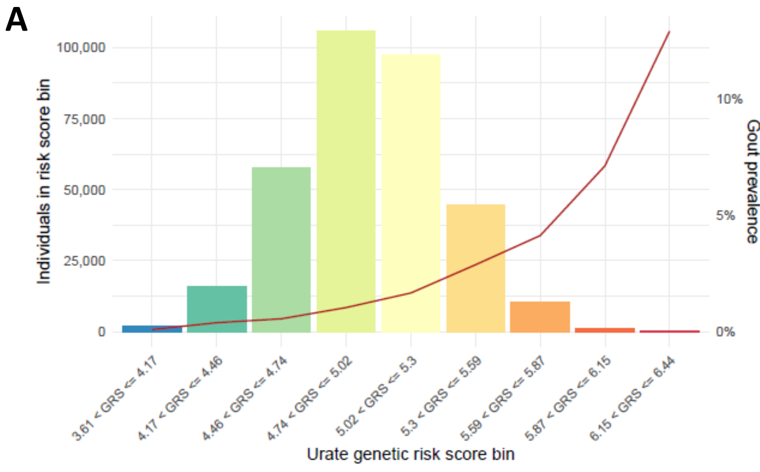
1590

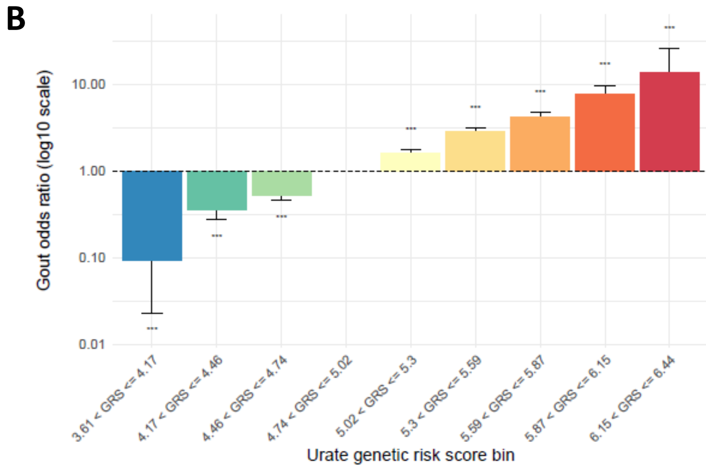
Figure 1

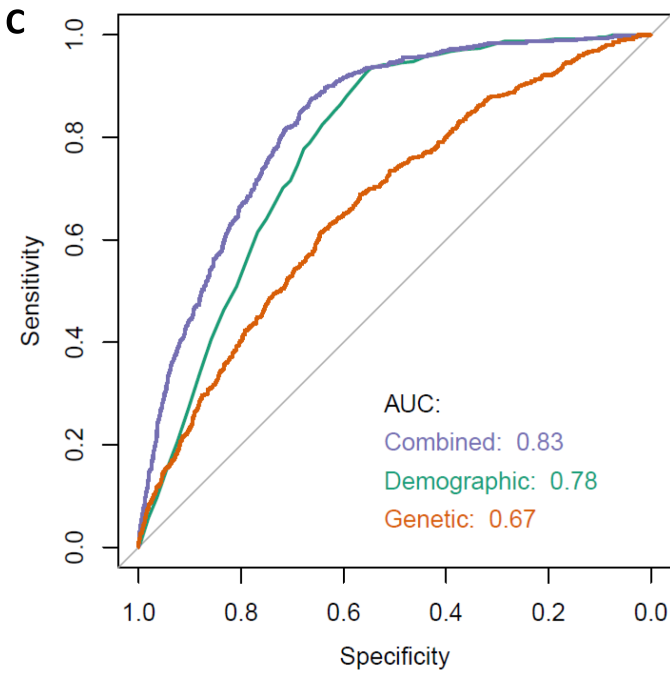


Track key
Gene Symbol
Nearest gene to
Index SNP
Novel
Previously reported
Index SNP Effect Size:
0.017
0.277
-log10(Purate)
-log10(Pgout)
Ancestry-associated
heterogeneity
Panc-het < 0.05/183
Pres-het < 0.05/183
Panc-het & Pres-het < 0.05/183
1
Chromosome

Figure 2







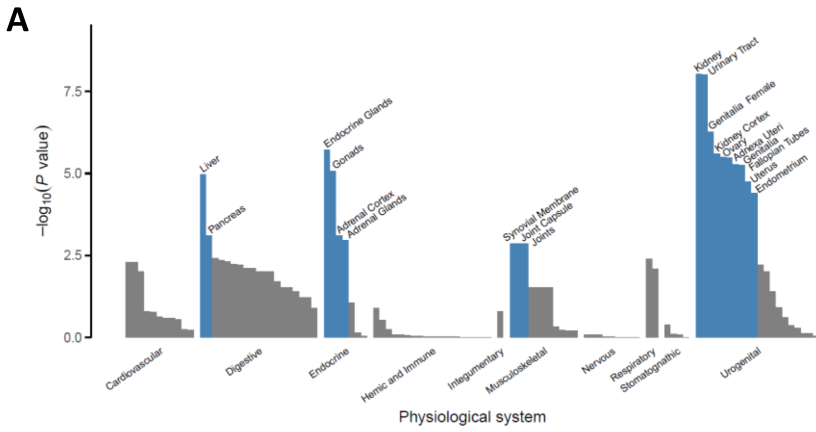
0.75
0.5
0.25
0

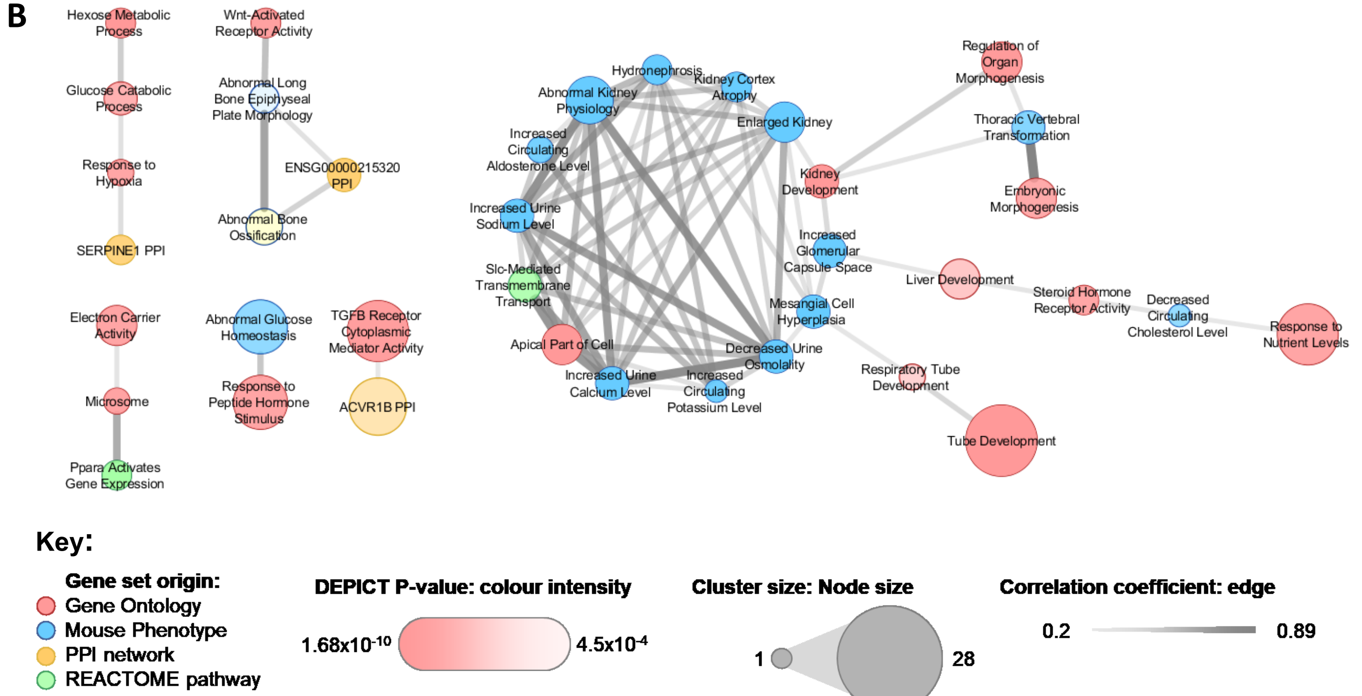
Gout
Alcohol and smoking
Amino acid, protein and lipid
Anthropometric
Cardiovascular and kidney
Education and behavior
Lung function, physical activity
Medication
Sleeping and others

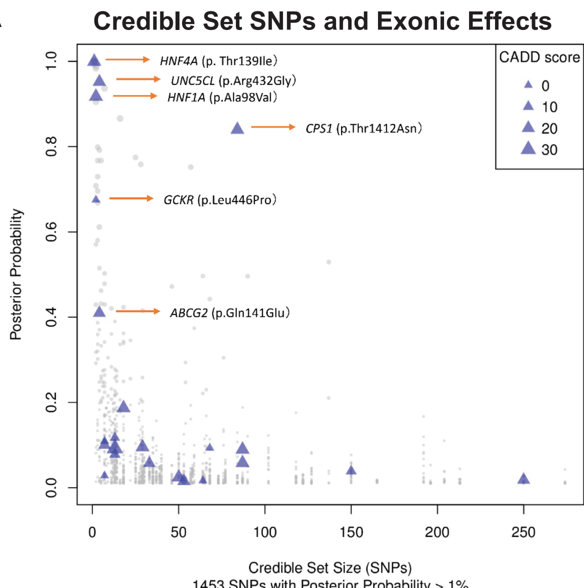
rg > 0
rg < 0

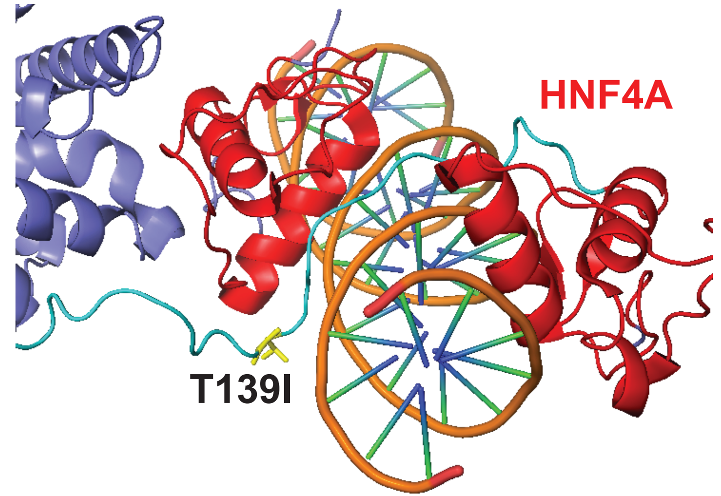
Self−reported gout
Why quit smoking: health precaution
Alcohol taken with meals
Weekly champagne intake
Weekly red wine intake
Maternal smoking around birth
Exposure to smoke outside home
Weekly spirits intake
Previously daily cigarettes
Alcohol intake frequency
Alcohol intake vs. 10 years ago
Pack years adult smoking
Pack years of smoking
Exposure to smoke at home
Free cholesterol in large HDL
Total cholesterol in large HDL
Cholesterol esters in large HDL
Phospholipids in large HDL
Total lipids in large HDL
Total cholesterol in HDL
Large HDL particles
Mean diameter for HDL particles
HDL cholesterol
Cholesterol esters in medium VLDL
Very large VLDL particles
Total lipids in very large VLDL
Total lipids in largest VLDL
Total cholesterol in medium VLDL
Free cholesterol in small VLDL
Triglycerides in very small VLDL
Cholesterol esters in large VLDL
Phospholipids in small VLDL
Total lipids in small VLDL
Total lipids in large VLDL
Large VLDL particles
Free cholesterol in large VLDL
Small VLDL particles
Glycoprotein acetyls
Mean diameter for VLDL particles
Phospholipids in chylomicrons/largest VLDL
Phospholipids in very large VLDL
Phospholipids in medium VLDL
Total lipids in very large VLDL
Triglycerides in large VLDL
Total cholesterol in large VLDL
Medium VLDL particles
Free cholesterol in medium VLDL
Phospholipids in large VLDL
Triglycerides in chylomicrons/largest VLDL
Self−reported high cholesterol
Isoleucine
Serum total triglycerides
Phospholipids in medium VLDL
Triglycerides in large VLDL
Triglycerides
Triglycerides in small VLDL
Triglycerides in medium VLDL
Triglycerides in small HDL
Impedance of whole body
Impedance of right arm
Impedance of left arm
Impedance of left leg
Impedance of right leg
Birth weight
Trunk predicted mass
Trunk fat−free mass
Comparative body size age 10
Whole body fat−free mass
Whole body water mass
Childhood obesity
Right arm fat−free mass
Right arm predicted mass
Right leg fat−free mass
Right leg predicted mass
Left arm predicted mass
Left arm fat−free mass
Left leg predicted mass
Left leg fat−free mass
Hip circumference
Waist circumference − UKBB
Weight
Waist−to−hip ratio
Trunk fat percentage
Trunk fat mass
Right leg fat percentage
Left leg fat percentage
Body fat percentage
Overweight
Hip circumference
Left arm fat percentage
Right arm fat percentage
Whole body fat mass
Obesity class 1
Body mass index
Obesity class 1
Right arm fat mass
Body mass index − UKBB
Right leg fat mass
Left leg fat mass
Left arm fat mass
Extreme bmi
Waist circumference − UKBB
Obesity class 3
Obesity class 2
Waist circumference
No siblings illness
Fathers age at death
No diagnosed heart problem
No mother illness
Parents age at death
Mothers age at death
Fathers age at death − UKBB
eGFR
eGFR non−DM
No father illness
Urine sodium
Diabetic disease father
Heart disease father
Heart disease mother
Fasting glucose main effect
Coronary artery disease
Hypertension mother
Systolic blood pressure
Hypertension father
Self−reported heart attack/MI
Diastolic blood pressure
Diagnosed heart attack
Heart disease siblings
Angina pectoris
HOMA−B
Diabetes mother
Chronic ischaemic heart disease
Diagnosed hypertension
Self−reported hypertension
Hypertension siblings
Diagnosed angina
Self−reported angina
Diabetes siblings
Type 2 diabetes
Self−reported diabetes
Diagnosed diabetes
Leptin, not adjBMI
Chronic kidney disease
Diabetes related eye disease
Self−reported type 2 diabetes
Fasting insulin main effect
HOMA−IR
Age completed education
Years of schooling 2013
Years of schooling (proxy)
Education: A levels qualifications
Years of schooling 2016
Education: college degree
Education: no qualifications
Time spent watching TV
Maximum workload during fitness test
Usual walking pace
Non−work transport: cycle
Work transport: cycle
Physical activity: strenuous sports
Stair climbing frequency
Physical activity: walking
Physical activity: heavy DIY
Best FVC
FVC − UKBB
FEV1
Best FEV1
Financial difficulties
Townsend deprivation index
Frequency of tiredness/lethargy
Currently unable to work
Knee pain
Wheeze or whistling in chest
Physical activity: none
Working week length main job
Serious illness, injury or assault
Overall health rating
Long−standing illness or disability
Basal metabolic rate
Health satisfaction
Leg pain on walking
Frequency travelling from home to job
Chest pain or discomfort walking
Shortness of breath walking
Number of self−reported illnesses
No medication in men
No medication in women
No pain relief medication
Paracetamol medication
Other prescription medications
Number of treatments/medications
Hypertensive medication in women
Hypertensive medication in men
Aspirin medication
Cholesterol medication in men
Cholesterol medication in women
Acceleration standard deviation
Relative age voice broke
Frequency of light DIY
Age at menarche − UKBB
Age at last live birth
Age at first live birth
Age at menarche
Relative age of first facial hair
Left heel bone mineral density
Snoring
Self−reported hypothyroidism
Schizophrenia
Nap during day
Pain in throat and chest
Education: CSEs
Abdominal and pelvic pain
Cholelithiasis

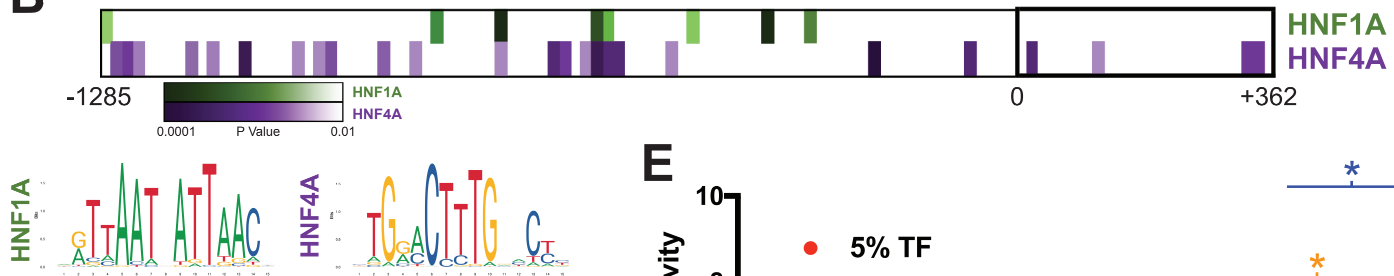
Figure 4

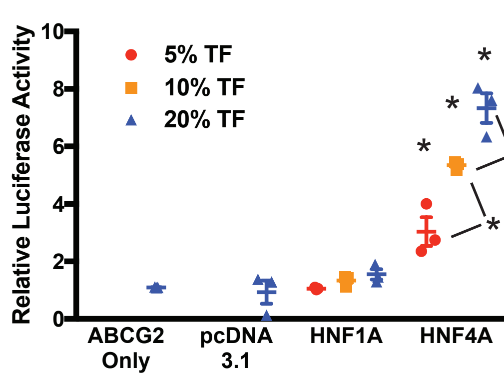


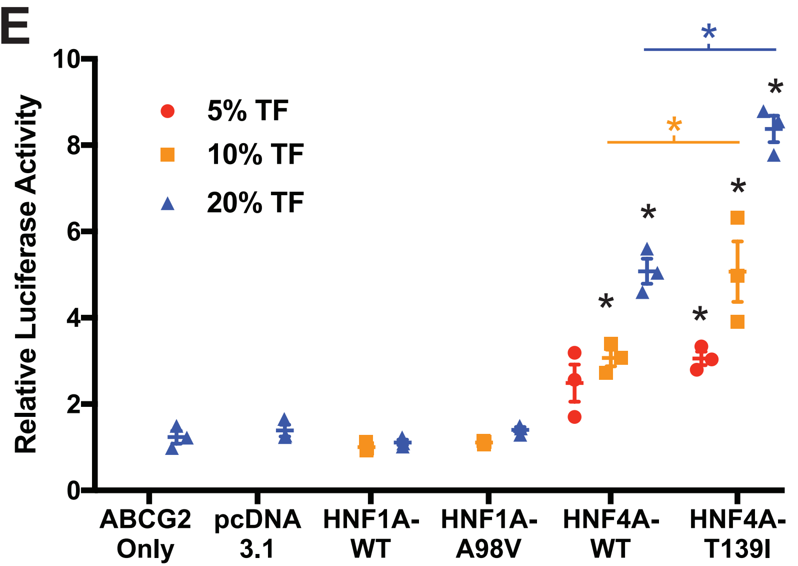



A
Credible Set SNPs and Exonic Effects
HNF4A (p. Thr139Ile)
UNC5CL (p.Arg432Gly)
HNF1A (p.Ala98Val)
CPS1 (p.Thr1412Asn)
GCKR (p.Leu446Pro)
ABCG2 (p.Gln141Glu)
CADD score
0
10
20
30
Posterior Probability
Credible Set Size (SNPs)
1453 SNPs with Posterior Probability > 1%


D
HNF4A
T139I


B
HNF1A
HNF4A
-1285
0
+362
P Value
0.0001
0.01

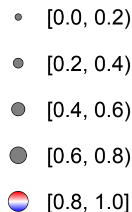

C
Relative Luciferase Activity
5% TF
10% TF
20% TF
ABCG2 Only
pcDNA 3.1
HNF1A
HNF4A


E
Relative Luciferase Activity
5% TF
10% TF
20% TF
ABCG2 Only
pcDNA 3.1
HNF1A-WT
HNF1A-A98V
HNF4A-WT
HNF4A-T139I

ARL6IP5, rs11128111
FGF5, rs10857147
CHRNA3, rs55781567
UMOD, rs4997081
GALNTL5, rs10480300
AAK1, rs12987661
NCOA7, rs4897160
BICC1, rs1649078
HLF, rs3794748
SFMBT1, rs2581817
MUC1, rs2070803
SLC16A9, rs1171617
SLC17A4, rs1359232

Locus
Renal Glomerular
Renal Tubulointerstitial
TCGA Kidney Cortex
Adipose Subcutaneous
Adrenal Gland
Artery Aorta
Artery Coronary
Artery Tibial
Brain Caudate Basal Ganglia
Brain Hippocampus
Brain Nucleus Accumbens Basal Ganglia
Breast Mammary Tissue
Colon Sigmoid
Colon Transverse
Esophagus Gastroesophageal Junction
Esophagus Mucosa
Esophagus Muscularis
Heart Atrial Appendage
Heart Left Ventricle
Liver
Lung
Nerve Tibial
Ovary
Pancreas
Skin Not Sun Exposed Suprapubic
Skin Sun Exposed Lower Leg
Spleen
Stomach
Testis
Thyroid

Posterior probability of colocalization



Change in gene expression with increased urate

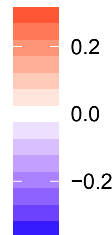


Table 1: Genes implicated as causal via identification of missense variants with high probability of driving the urate association signal. Genes are included if they contain a missense variant with posterior probability of association of >50% or mapping into a small credible set (≤5 variants).

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, DNase-I hypersensitivity site; CADD, Combined Annotation Dependent Depletion phred score; EA, European ancestry.