

## GWAS defines pathogenic signaling pathways and prioritizes drug targets for IgA nephropathy

Krzysztof Kiryluk<sup>1,2^</sup>, Elena Sanchez-Rodriguez<sup>1\*</sup>, Xu-jie Zhou<sup>3\*</sup>, Francesca Zanoni<sup>1</sup>, Lili Liu<sup>1</sup>, Nikol Mladkova<sup>1</sup>, Atlas Khan<sup>1</sup>, Maddalena Marasa<sup>1</sup>, Jun Y. Zhang<sup>1</sup>, Olivia Balderes<sup>1</sup>, Simone Sanna-Cherchi<sup>1,2</sup>, Andrew S. Bomback<sup>1</sup>, Pietro A. Canetta<sup>1</sup>, Gerald B. Appel<sup>1</sup>, Jai Radhakrishnan<sup>1</sup>, Hernan Trimarchi<sup>4</sup>, Ben Sprangers<sup>5,6</sup>, Daniel C. Cattran<sup>7</sup>, Heather Reich<sup>7</sup>, York Pei<sup>7</sup>, Pietro Ravani<sup>8</sup>, Kresimir Galesic<sup>9</sup>, Dita Maixnerova<sup>10</sup>, Vladimir Tesar<sup>10</sup>, Benedicte Stengel<sup>11</sup>, Marie Metzger<sup>11</sup>, Guillaume Canaud<sup>12</sup>, Nicolas Maillard<sup>13</sup>, Francois Berthoux<sup>13</sup>, Laureline Berthelot<sup>14</sup>, Evangeline Pillebout<sup>15</sup>, Renato Monteiro<sup>15</sup>, Raoul Nelson<sup>16</sup>, Robert Wyatt<sup>17,18</sup>, William Smoyer<sup>19</sup>, John Mahan<sup>19</sup>, Al-Akash Samhar<sup>20</sup>, Guillermo Hidalgo<sup>21</sup>, Alejandro Quiroga<sup>22</sup>, Patricia Weng<sup>23</sup>, Raji Sreedharan<sup>24</sup>, David Selewski<sup>25</sup>, Keefe Davis<sup>26</sup>, Mahmoud Kallash<sup>27</sup>, Tetyana L. Vasylyeva<sup>28</sup>, Michelle Rheault<sup>29</sup>, Aftab Chishtii<sup>30</sup>, Daniel Ranch<sup>31</sup>, Scott E. Wenderfer<sup>32</sup>, Dmitry Samsonov<sup>33</sup>, Donna J. Claes<sup>34</sup>, Akchurin Oleh<sup>35</sup>, Dimitrios Goumenos<sup>36</sup>, Maria Stangou<sup>36</sup>, Judit Nagy<sup>37</sup>, Tibor Kovacs<sup>37</sup>, Enrico Fiaccadori<sup>38</sup>, Antonio Amoroso<sup>39</sup>, Cristina Barlassina<sup>40</sup>, Daniele Cusi<sup>40</sup>, Lucia Del Vecchio<sup>41</sup>, Giovanni Giorgio Battaglia<sup>42</sup>, Monica Bodria<sup>43</sup>, Emanuela Boer<sup>44</sup>, Luisa Bono<sup>45</sup>, Giuliano Boscutti<sup>46</sup>, Gianluca Caridi<sup>47</sup>, Francesca Lugani<sup>48</sup>, GianMarco Ghiggeri<sup>47</sup>, Rosanna Coppo<sup>48</sup>, Licia Peruzzi<sup>48</sup>, Vittoria Esposito<sup>49</sup>, Ciro Esposito<sup>49</sup>, Sandro Feriozzi<sup>50</sup>, Rosaria Polci<sup>50</sup>, Giovanni Frasca<sup>51</sup>, Marco Galliani<sup>52</sup>, Maurizio Garozzo<sup>53</sup>, Adele Mitrotti<sup>54</sup>, Loreto Gesualdo<sup>54</sup>, Simona Granata<sup>55</sup>, Gianluigi Zaza<sup>55</sup>, Francesco Londrino<sup>56</sup>, Riccardo Magistroni<sup>57</sup>, Isabella Pisani<sup>58</sup>, Andrea Magnano<sup>58</sup>, Carmelita Marcantoni<sup>59</sup>, Piergiorgio Messa<sup>60</sup>, Renzo Mignani<sup>61</sup>, Antonello Pani<sup>62</sup>, Claudio Ponticelli<sup>63</sup>, Dario Roccatello<sup>64</sup>, Maurizio Salvadori<sup>65</sup>, Erica Salvi<sup>66</sup>, Domenico Santoro<sup>67</sup>, Guido Gembillo<sup>67</sup>, Silvana Savoldi<sup>68</sup>, Donatella Spotti<sup>69</sup>, Pasquale Zamboli<sup>70</sup>, Claudia Izzi<sup>71</sup>, Federico Alberici<sup>71</sup>, Elisa Delbarba<sup>71</sup>, Michał Florczak<sup>72</sup>, Natalia Krata<sup>72</sup>, Krzysztof Mucha<sup>72,73</sup>, Leszek Pączek<sup>72,73</sup>, Stanisław Niemczyk<sup>74</sup>, Barbara Moszczuk<sup>72,75</sup>, Malgorzata Pańczyk-Tomaszewska<sup>76</sup>, Malgorzata Mizerska-Wasiak<sup>76</sup>, Agnieszka Perkowska-Ptasińska<sup>77</sup>, Teresa Bączkowska<sup>78</sup>, Magdalena Durlik<sup>78</sup>, Krzysztof Pawlaczyk<sup>79</sup>, Przemysław Sikora<sup>80</sup>, Marcin Zaniew<sup>81</sup>, Dorota Kaminska<sup>82</sup>, Magdalena Krajewska<sup>82</sup>, Izabella Kuzmiuk-Glembin<sup>83</sup>, Zbigniew Heleniak<sup>83</sup>, Barbara Bullo-Piontecka<sup>83</sup>, Tomasz Liberek<sup>83</sup>, Alicja Dębska-Slizien<sup>83</sup>, Tomasz Hryszko<sup>84</sup>, Anna Materna-Kiryluk<sup>85</sup>, Monika Miklaszewska<sup>86</sup>, Maria Szczepańska<sup>87</sup>, Katarzyna Dyga<sup>87</sup>, Edyta Machura<sup>87</sup>, Katarzyna Siniewicz-Luzeńczyk<sup>88</sup>, Monika Pawlak-Bratkowska<sup>88</sup>, Marcin Tkaczyk<sup>88</sup>, Dariusz Runowski<sup>89</sup>, Norbert Kwellla<sup>90</sup>, Dorota Drożdż<sup>91</sup>, Ireneusz Habura<sup>92</sup>, Florian Kronenberg<sup>93</sup>, Larisa Prikhodina<sup>94</sup>, David van Heel<sup>95</sup>, Bertrand Fontaine<sup>96,97</sup>, Chris Cotsapas<sup>98</sup>, Cisca Wijmenga<sup>99</sup>, Andre Franke<sup>100</sup>, Vito Annese<sup>101</sup>, Peter K. Gregersen<sup>102</sup>, Sreeja Parameswaran<sup>103</sup>, Matthew Weirauch<sup>103,104</sup>, Leah Kottyan<sup>103,104</sup>, John B Harley<sup>105</sup>, Hitoshi Suzuki<sup>106</sup>, Ichiei Narita<sup>107</sup>, Shin Goto<sup>107</sup>, Hajeong Lee<sup>108</sup>, Dong Ki Kim<sup>108</sup>, Yon Su Kim<sup>109</sup>, Jin-Ho Park<sup>110</sup>, BeLong Cho<sup>110,111</sup>, Murim Choi<sup>109</sup>, Ans Van Wijk<sup>112</sup>, Ana Huerta<sup>113</sup>, Elisabet Ars<sup>114</sup>, Jose Ballarin<sup>114</sup>, Sigrid Lundberg<sup>115</sup>, Bruno Vogt<sup>116</sup>, Laila-Yasmin Mani<sup>116</sup>, Yasar Caliskan<sup>117</sup>, Jonathan Barratt<sup>118</sup>, Thilini Abeygunaratne<sup>119</sup>, Philip A. Kalra<sup>119</sup>, Daniel P. Gale<sup>120</sup>, Ulf Panzer<sup>121</sup>, Thomas Rauen<sup>122</sup>, Jürgen Floege<sup>122</sup>, Pascal Schlosser<sup>123</sup>, Arif B. Ekici<sup>124</sup>, Kai-Uwe Eckardt<sup>125,126</sup>, Nan Chen<sup>127</sup>, Jingyuan Xie<sup>127</sup>, Richard P. Lifton<sup>128,129</sup>, Ruth J. F. Loos<sup>130,131</sup>, Eimear E. Kenny<sup>130,132,133</sup>, Iuliana Ionita-Laza<sup>134</sup>, Anna Köttgen<sup>123</sup>, Bruce Julian<sup>135</sup>, Jan Novak<sup>135</sup>, Francesco Scolari<sup>71</sup>, Hong Zhang<sup>3</sup>, Ali G. Gharavi<sup>1,2^</sup>.

### Affiliations:

1. Department of Medicine, Vagelos College of Physicians & Surgeons, Columbia University, New York, New York, USA
2. Institute for Genomic Medicine, Columbia University, New York, New York, USA
3. Renal Division, Peking University First Hospital, Peking University Institute of Nephrology, Beijing, China
4. Nephrology Service, Hospital Británico de Buenos Aires, Perdriel 74 (1280), Buenos Aires, Argentina.
5. Department of Microbiology and Immunology, Laboratory of Molecular Immunology, KU Leuven, Leuven, Belgium
6. Division of Nephrology, University Hospitals Leuven, Leuven, Belgium
7. Department of Nephrology, University of Toronto, Toronto General Hospital, Toronto, Ontario, Canada.
8. Division of Nephrology, Department of Internal Medicine, University of Calgary, Calgary, Canada
9. Zagreb Medical School, Zagreb, Croatia
10. 1st Faculty of Medicine and General University Hospital, Charles University, Prague, Czech Republic
11. Centre for Research in Epidemiology and Population Health (CESP), Paris-Saclay University, Versailles Saint Quentin University, Inserm, Clinical Epidemiology Team, Villejuif, France
12. Université de Paris, Hôpital Necker-Enfants Malades, INSERM U1151, Paris, France
13. Nephrology, Dialysis, and Renal Transplantation Department, University North Hospital, Saint Etienne, France
14. CRTI UMR1064, Inserm, Nantes University, Nantes, France
15. Center for Research on Inflammation, University of Paris, Inserm U1149 & Cnrs ERL8252, Paris, France
16. Division of Pediatric Nephrology, Department of Pediatrics, University of Utah, USA
17. Division of Pediatric Nephrology, University of Tennessee Health Sciences Center, Memphis, Tennessee, USA
18. Children's Foundation Research Center, Le Bonheur Children's Hospital, Memphis, Tennessee, USA

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

19. Division of Pediatric Nephrology, Nationwide Children's Hospital, USA
20. Division of Pediatric Nephrology, Driscoll Children's Hospital, USA
21. Division of Pediatric Nephrology, Department of Pediatrics, HMH Hackensack University Medical Center, New Jersey, USA
22. Division of Pediatric Nephrology, Mattel Children's Hospital, USA
23. Division of Pediatric Nephrology, Medical College of Wisconsin, USA
24. Division of Pediatric Nephrology, Mott Children's Hospital, USA
25. Division of Pediatric Nephrology, Department of Pediatrics, the Medical University of South Carolina (MUSC), South Carolina, USA
26. Division of Pediatric Nephrology, SUNY Buffalo, USA
27. Division of Pediatric Nephrology, Department of Pediatrics, Nationwide Children's Hospital, OH, USA
28. Division of Pediatric Nephrology, University of Minnesota, USA
29. Division of Pediatric Nephrology, University of Kentucky, USA
30. Division of Pediatric Nephrology, Department of Pediatrics, University of Kentucky, USA
31. Division of Pediatric Nephrology, Baylor College of Medicine/Texas Children's Hospital, USA
32. Division of Pediatric Nephrology, Boston Children's Hospital, USA
33. Division of Pediatric Nephrology, Department of Pediatrics, New York Medical College, NY, USA
34. Division of Pediatric Nephrology, Cornell University, USA
35. The Aristotle University of Thessaloniki, Greece
36. 2nd Department of Internal Medicine, Nephrological and Diabetological Center, University of Pécs, Hungary
37. Department of Medicine and Surgery, University of Parma, Parma, Italy.
38. Department of Medical Sciences, University of Turin, Italy
39. Renal Division, Dipartimento di Medicina, Chirurgia e Odontoiatria, San Paolo Hospital, School of Medicine, University of Milan, Milan, Italy
40. Department of Nephrology and dialysis, ASST Lariana, Como, Italy
41. Unità Operativa di Nefrologia e Dialisi, Ospedale di Acireale, Acireale, Italy
42. University of Messina, Italy
43. Division of Nephrology and Dialysis, Gorizia Hospital, Gorizia, Italy
44. Nephrology and Dialysis, A.R.N.A.S. Civico and Benfratellio, Palermo, Italy
45. Nephrology, Dialysis and Renal Transplant Unit, S. Maria della Misericordia Hospital, ASUFC Udine, Italy
46. Division of Nephrology, Dialysis and Transplantation, IRCCS Giannina Gaslini Institute, Genova, Italy
47. Regina Margherita Children's Hospital, Torino, Italy
48. ICS Maugeri SpA, University of Pavia, Italy
49. Belcolle Hospital, Viterbo, Italy
50. Division of Nephrology, Dialysis and Renal Transplantation, Riuniti Hospital, Ancona, Italy
51. Sandro Pertini Hospital, Rome, Italy;
52. Unità Operativa di Nefrologia e Dialisi, Ospedale di Acireale, Acireale, Italy
53. Nephrology, Dialysis and Transplantation Unit, Department of Emergency and Organ Transplantation, University of Bari Aldo Moro, Bari, Italy
54. Renal Unit, Department of Medicine, University of Verona, Italy
55. S. Andrea Hospital, La Spezia, Italy
56. Department of Surgical, Medical, Dental, Oncologic and Regenerative Medicine, University of Modena and Reggio Emilia
57. Department of Medicine and Surgery, University of Parma, Italy
58. Nephrology Division, San Marco Hospital, Catania, Italy
59. Nephrology Dialysis and Kidney Transplant Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Università degli studi di Milano, Milan, Italy
60. Azienda Unità Sanitaria Locale Rimini, Italy
61. Department of Nephrology and Dialysis, G. Brotzu Hospital, Cagliari, Italy
62. Nephrology and Dialysis, Ospedale Maggiore, Milano, Italy (retired)
63. Nephrology and Dialysis Unit, G. Bosco Hub Hospital (ERK-net Member) and University of Torino, Italy
64. Division of Nephrology and Renal Transplantation, Carreggi Hospital, Florence, Italy
65. Renal Division, DMCO (Dipartimento di Medicina, Chirurgia e Odontoiatria), San Paolo Hospital, School of Medicine, University of Milan, Milan, Italy
66. Unit of Nephrology and Dialysis, AOU G Martino, University of Messina, Italy
67. Unit of Nephrology and Dialysis ASL TO4, Cirié, Turin, Italy
68. San Raffaele Hospital, Milan, Italy
69. NephroCare Italy, Naples, Italy
70. Department of Medical and Surgical Specialties and Nephrology Unit, University of Brescia – ASST Spedali Civili, Brescia, Italy
71. Department of Immunology, Transplantology and Internal Diseases, Medical University of Warsaw, Poland
72. Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland
73. Department of Internal Disease, Nephrology and Dialysis, Military Institute of Medicine, Warsaw, Poland
74. Department of Clinical Immunology, Medical University of Warsaw, Warsaw, Poland
75. Department of Pediatrics and Nephrology, Medical University of Warsaw, Warsaw, Poland
76. Department of Pathology, Medical University of Warsaw, Poland
77. Department of Transplantation Medicine, Nephrology and Internal Diseases, Medical University of Warsaw, Poland
78. Department of Nephrology, Transplantology and Internal Medicine, Poznan Medical University, Poznan, Poland
79. Department of Pediatric Nephrology Medical University of Lublin, Poland
80. Department of Pediatrics, University of Zielona Góra, Zielona Góra, Poland
81. Clinical Department of Nephrology and Transplantation Medicine, Wrocław Medical University, Wrocław, Poland
82. Department of Nephrology, Transplantology and Internal Diseases, Medical University of Gdansk, Poland
83. 2nd Department of Nephrology and Hypertension with Dialysis Unit, Medical University of Białystok, Poland
84. Department of Genetics, Medical University of Poznan, Poznan, Poland
85. Department of Pediatric Nephrology and Hypertension, Jagiellonian University Medical College, Krakow, Poland
86. Department of Pediatrics, Faculty of Medical Sciences in Zabrze, Medical University of Silesia in Katowice, Poland
87. Department of Pediatrics, Immunology and Nephrology, Polish Mother's Memorial Hospital Research Institute, Lodz, Poland
88. Department of Nephrology, Kidney Transplantation and Hypertension, Children's Memorial Health Institute, Warsaw, Poland
89. Department of Nephrology, Hypertension and Internal Medicine, University of Warmia and Mazury in Olsztyn, Olsztyn, Poland
90. Department of Pediatric Nephrology and Hypertension, Jagiellonian University Medical College, Krakow
91. Department of Nephrology, Karol Marcinkowski Hospital, Zielona Góra, Poland
92. Institute of Genetic Epidemiology, Department of Genetics and Pharmacology, Medical University of Innsbruck, Innsbruck, Austria.

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

93. Division of Inherited & Acquired Kidney Diseases, Veltishev Research and Clinical Institute for Pediatrics of the Pirogov Russian National Research Medical University, Moscow, Russia
94. Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK
95. Sorbonne University, INSERM. Center of Research in Myology - UMR 974, Institute of Myology, University Hospital Pitie-Salpetriere, Paris, France
96. Assistance Publique-Hôpitaux de Paris (AP-HP), Service of Neuro-Myology, University Hospital Pitie-Salpetriere, Paris, France
97. Departments of Neurology and Genetics, Yale University
98. University of Groningen, Groningen, the Netherlands
99. Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany
100. CBP American Hospital, Dubai, United Arab Emirates
101. Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, North Shore LIJ Health System, New York, USA
102. Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA
103. University of Cincinnati College of Medicine, Cincinnati, Ohio, USA
104. US Department of Veterans Affairs Medical Center, Cincinnati, OH and Cincinnati Educational and Research for Veterans Foundation, Cincinnati, Ohio, USA
105. Department of Nephrology, Juntendo University Faculty of Medicine, Tokyo, Japan
106. Division of Clinical Nephrology and Rheumatology, Kidney Research Center, Niigata University Graduate School of Medical and Dental Sciences.
107. Internal Medicine, Seoul National University College of Medicine, Seoul, Korea
108. Biomedical Science, Seoul National University College of Medicine, Seoul, Korea
109. Department of Family Medicine, Seoul National University College of Medicine and Seoul National University Hospital, Seoul, Korea
110. Institute on Aging, Seoul National University College of Medicine, Seoul, Korea
111. Amsterdam University Medical Centre, VU University Medical Center (VUMC) location, Amsterdam, Netherlands
112. Hospital Universitario Puerta del Hierro Majadahonda, Spain. REDINREN, IISCIH, Spain.
113. Molecular Biology Laboratory and Nephrology Department, Fundació Puigvert, Instituto de Investigaciones Biomédicas Sant Pau, Universitat Autònoma de Barcelona, REDINREN, IISCIH, Barcelona, Spain
114. Department of Nephrology, Danderyd University Hospital, and Department of Clinical Sciences, Karolinska Institutet, Stockholm, Sweden
115. Department of Nephrology and Hypertension, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland
116. Division of Nephrology, Saint Louis University, Saint Louis, MO, USA
117. John Walls Renal Unit, University Hospitals of Leicester, Leicester, UK
118. National Institute of Health Research, UK
119. Salford Royal Hospital, Salford, UK
120. University of Hamburg, Germany
121. Division of Nephrology and Clinical Immunology, RWTH Aachen University, Aachen, Germany
122. Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany.
123. Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany.
124. Department of Nephrology and Medical Intensive Care, Charité-Universitätsmedizin Berlin, Berlin, Germany.
125. Department of Nephrology and Hypertension, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany.
126. Department of Nephrology, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China
127. Department of Genetics, Yale School of Medicine, New Haven, Connecticut, USA
128. Laboratory of Human Genetics and Genomics, The Rockefeller University, New York, New York, USA
129. The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA
130. Novo Nordisk Foundation Center for Basic Metabolic Research, Department of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
131. Department of Genetics and Genomic Sciences, Mount Sinai Health System, New York, New York, USA
132. Center for Population Genomic Health, Icahn School of Medicine at Mount Sinai, New York, New York, USA
133. Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, USA
134. Departments of Microbiology and Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA

**\*\* Equal contributions**

**^ Corresponding Authors:**

Krzysztof Kiryluk, M.D. M.S.  
Division of Nephrology, Department of Medicine  
Institute for Genomic Medicine  
College of Physicians & Surgeons, Columbia University  
1150 St. Nicholas Ave., Russ Berrie Pavilion 412E, New York, NY 10032  
[kk473@columbia.edu](mailto:kk473@columbia.edu)

Ali G. Gharavi, M.D.  
Division of Nephrology, Department of Medicine  
Institute for Genomic Medicine  
College of Physicians & Surgeons, Columbia University  
1150 St. Nicholas Ave., Russ Berrie Pavilion 413, New York, NY 10032  
[ag2239@cumc.columbia.edu](mailto:ag2239@cumc.columbia.edu)

## 2 ABSTRACT

3

4 IgA nephropathy (IgAN) is a progressive form of kidney disease defined by glomerular deposition of IgA.

5 We performed a genome-wide association study involving 10,146 kidney biopsy-diagnosed IgAN cases

6 and 28,751 matched controls across 17 international cohorts. We defined 30 independent genome-wide

7 significant risk loci jointly explaining 11% of disease risk. A total of 16 loci were novel, including

8 *TNFSF4*, *REL*, *CD28*, *CXCL8/PF4V1*, *LY86*, *LYN*, *ANXA3*, *TNFSF8/15*, *REEP3*, *ZMIZ1*, *RELA*, *ETS1*,

9 *IGH*, *IRF8*, *TNFRSF13B* and *FCAR*. The SNP-based heritability of IgAN was estimated at 23%. We

10 observed a positive genetic correlation between IgAN and total serum IgA levels, allergy, tonsillectomy,

11 and several infections, and a negative correlation with inflammatory bowel disease. All significant non-

12 HLA loci shared with serum IgA levels had a concordant effect on the risk of IgAN. Moreover, IgAN loci

13 were globally enriched in gene orthologs causing abnormal IgA levels when genetically manipulated in

14 mice. The explained heritability was enriched in the regulatory elements of cells from the immune and

15 hematopoietic systems and intestinal mucosa, providing support for the pathogenic role of extra-renal

16 tissues. The polygenic risk of IgAN was associated with early disease onset, increased lifetime risk of

17 kidney failure, as well as hematuria and several other traits in a phenome-wide association study of

18 590,515 individuals. In the comprehensive functional annotation analysis of candidate causal genes across

19 genome-wide significant loci, we observed the convergence of biological candidates on a common set of

20 inflammatory signaling pathways and cytokine ligand-receptor pairs, prioritizing potential new drug

21 targets.

22

23

24

25

26

## 2 INTRODUCTION

3

4 IgA nephropathy (IgAN) is a common form of immune-mediated glomerulonephritis characterized by  
5 glomerular deposition of IgA-containing immune complexes and manifesting with hematuria, proteinuria,  
6 and often kidney failure. Examination of kidney-biopsy tissue and demonstration of glomerular IgA  
7 deposits is required to establish the diagnosis. No approved targeted therapies presently exist for IgAN,  
8 and a large fraction of cases progress to kidney failure requiring kidney transplantation or dialysis. There  
9 are no validated molecular predictors of progression to kidney failure.

10

11 As the diagnosis requires a kidney biopsy, genetic discoveries have been hindered by small sample sizes  
12 of the existing genetic cohorts. Approximately 15 independent loci have been previously identified in  
13 association with IgAN, implicating defects in the complement pathway, intestinal network of IgA  
14 production, and innate immunity against mucosal pathogens<sup>1-7</sup>. These findings have already led to the  
15 reformulation of the existing IgAN pathogenesis model, with most candidate mechanisms mapping to the  
16 immune system rather than the kidney<sup>8,9</sup>. Nevertheless, prior GWAS had a two-stage design, with sample  
17 size of the discovery stage often limiting the power to discover new loci.

18

19 Herein, we report a well-powered GWAS discovery study for IgAN involving 38,897 individuals (10,146  
20 kidney biopsy-defined cases and 28,751 controls) recruited across 17 international cohorts. With the  
21 discovery of 16 novel non-HLA GWAS loci, we provide strong support for a highly polygenic  
22 architecture of IgAN. We assess functional consequences of the risk alleles, define causal cell types and  
23 signaling pathways, and explore genetic correlations and pleiotropic associations of the risk loci. We  
24 demonstrate that IgAN polygenic score predicts kidney disease outcomes. Importantly, we report  
25 convergence of multiple risk loci on the set of common signaling pathways and ligand-receptor pairs  
26 involved in the regulation of IgA production, prioritizing plausible new molecular drug targets.

2

## 3 **RESULTS**

4

### 5 **Study Design**

6

7 We performed a standardized GWAS and meta-analysis of 17 independent international case-control  
8 cohorts (12 newly genotyped and five previously published GWAS cohorts) comprising a total of 38,897  
9 individuals (10,146 biopsy-proven cases and 28,751 controls). Of the 17 cohorts, 11 were of European  
10 ancestry with participants recruited from nephrology centers in Italy, Poland, Germany, France, Belgium,  
11 Czech Republic, Hungary, Croatia, Turkey, Spain, Sweden, U.K., U.S., Canada, and Argentina, and six  
12 cohorts were of East-Asian ancestry with participants recruited from nephrology centers in China, Japan,  
13 and Korea. A total of 14 cohorts (8,139 cases and 17,713 controls) were genotyped with high-density SNP  
14 arrays, ancestrally matched using principal component-based methods, and imputed using whole genome  
15 sequence reference panels specific for each ethnicity, and three additional cohorts (2,007 cases and 11,038  
16 controls) were genotyped with the ImmunoChip platform (**Supplementary Table 1**).

17

18 Detailed description of each cohort, genotyping platform, quality control analysis, and ancestry and  
19 imputation analyses are provided as **Supplementary Notes**. Our primary discovery involved the  
20 combined analysis of all 17 cohorts under a log-additive genetic model. Additional exploratory analyses  
21 were conducted to identify any potential ancestry-specific and sex-specific locus, including under  
22 alternative (dominant and recessive) genetic models.

23

### 24 **Genome-wide significant loci**

25

26 The results of combined meta-analyses across all cohorts are summarized in **Figure 1, Tables 2 and**  
27 **Supplementary Tables 2-4**. We confirmed multiple independently genome-wide significant ( $p < 5 \times 10^{-8}$ )

2 signals in the HLA region, with an overall  $\lambda=1.048$  and 1.042 before and after excluding the HLA region  
3 (**Extended Data Figure 1**). In addition, we detected 24 independently associated non-HLA loci at a  
4 genome-wide significance, including eight known non-HLA loci (*CFH*, *DEFA1/4*, *CARD9*,  
5 *ITGAM/ITGAX*, *TNFSF13*, *LIF/OSM*, *FCRL3*, *IRF4/DUSP22*), 16 novel loci (*TNFSF4*, *CD28*, *REL*,  
6 *PF4V1/CXCL1*, *LY86/RREB1*, *LYN*, *ANXA3*, *TNFSF8/15*, *REEP3*, *ZMIZ1/PPIF*, *OVOL1/RELA*, *ETS1*,  
7 *IGH*, *IRF8*, *TNFRSF13B* and *FCAR*, see **Extended Data Figure 2**), in addition to 48 independent  
8 suggestive non-HLA signals at  $P < 1 \times 10^{-5}$  (**Supplementary Tables 2 and 3**). Ethnicity-specific meta-  
9 analyses revealed another independent genome-wide significant association that was evident only in the  
10 East-Asian cohorts (*CCR6* locus, **Extended Data Figure 3a**) and 11 additional suggestive signals in this  
11 ancestral group (**Supplementary Table 4**). One of the suggestive signals in the combined meta-analysis  
12 (*PADI3/PADI4* locus) reached genome-wide significance in the East-Asian meta-analysis under a  
13 recessive model (**Extended Data Figure 3b**). The European-only meta-analysis showed a total of 19  
14 suggestive signals, but no additional new genome-wide significant loci (**Supplementary Table 4**). No  
15 gender-specific loci were found in a gender-stratified meta-analysis, or in the analysis of sex  
16 chromosomal markers.

17  
18 We next performed stepwise conditional analyses of the 24 genome-wide significant non-HLA loci, but  
19 only the *CFH* locus showed evidence of at least two independent genome-wide significant variants  
20 (**Supplementary Table 5 and Extended Data Figure 4a**). Stepwise conditional analyses of the *HLA*  
21 region revealed a complex pattern of association, with at least five independently genome-wide significant  
22 SNPs in the combined analysis of all cohorts (**Extended Data Figure 4b, Supplementary Table 6**). In  
23 addition, the patterns of association across the HLA region differed when stepwise conditioning was  
24 performed separately in East-Asian and European cohorts – a total of five independent signals were  
25 detected in the European cohorts, and four in the East-Asian cohorts. In the overall meta-analysis, we

2 observed an inverse relationship between minor allelic frequency of the top independently associated  
3 variants and their effect sizes, consistent with the effects of purifying selection (**Figure 1b**).  
4 Based on genome-wide summary statistics, we estimated the SNP-based heritability of IgAN at 0.23  
5 (95%CI: 0.15-0.30). Excluding the MHC region reduced SNP-based heritability estimate to 0.12 (95%CI:  
6 0.10-0.13), suggesting that HLA and non-HLA loci respectively contribute approximately 50% of the  
7 polygenic risk. Using the genetic risk score (expressed as the sum of the risk alleles weighted by their  
8 individual effect sizes) based on 30 independently genome-wide significant SNPs (30-SNP GRS)  
9 explained 11% of overall disease variance, a significant improvement compared to the 6% explained by  
10 the previous 15-SNP GRS<sup>4</sup>.

11

## 12 **Classical HLA alleles**

13

14 To better understand the signal at the HLA locus, we imputed amino-acid sequences and classical HLA  
15 alleles at four-digit resolution at class II (*HLA-DQB1*, *-DQA1* and *-DRB1*) and class I (*HLA-A*, *-B*, and *-C*)  
16 genes (see **Methods**). We used ethnicity-specific reference panels for imputation, followed by association  
17 testing within each ancestral group separately. The analysis of imputed amino-acid sequences in class II  
18 genes pointed to *DRB1* as the gene with most strongly associated polymorphic positions in both ancestral  
19 groups (**Extended Data Figure 5**). In East Asian cohorts, stepwise conditioning of multiallelic sites  
20 demonstrated independently significant associations at DRβ1 positions 11 and 71, and the same positions  
21 were also strongly associated with the disease in Europeans. Specifically, Proline at position 11 (in LD  
22 with Arginine at position 71 and corresponding to DRB1\*1501) conveyed significant protection in both  
23 ancestral groups (**Supplementary Table 7**). There was also an independently significant risk effect of  
24 Arginine at position 71 (in LD with Valine at position 11) with consistent direction of effect across both  
25 ancestries. In Europeans, we additionally observed significant effects of two substitutions at position 11,

2 Glycine (protective) and Leucine (risk). These two substitutions are less frequent in East Asian  
3 populations.

4

5 The association patterns of classical HLA alleles were complex, but generally consistent with the analysis  
6 of amino acid sequences. In East Asians, we observed a protective effect of the *DRB1\*1501-DQA1\*0102-*  
7 *DQB1\*0602* haplotype (DR15 serotype), and an independent risk effect of *DRB1\*0405*, with no  
8 significant associations after conditioning for both *DRB1\*1501* and *DRB1\*0405* (**Supplementary Table**  
9 **8**). In Europeans, we confirmed a strong protective association of the *DRB1\*1501-DQA1\*0102-*  
10 *DQB1\*0602* haplotype. *DRB1\*0405* had low allelic frequency in Europeans, thus this association was not  
11 replicated. Instead, we observed three additional independent European haplotypes (rare in East Asians),  
12 including two protective haplotypes, *DRB1\*0301-DQA1\*0501-DQB1\*0201* (DR3 serotype) and  
13 *DRB1\*0701-DQA1\*0201-DQB1\*0202/0203* (DR7 serotype), and one risk haplotype *DRB1\*0101-*  
14 *DQA1\*0101-DQB1\*0501* (DR1 serotype, **Supplementary Table 9**). After conditioning for the four  
15 independently significant *DRB1* alleles residing on these haplotypes, we observed additional independent  
16 protective associations of *DQA1\*0102* and *DPA1\*0103* in Europeans. There were no independently  
17 significant associations for the class I genes. Collectively, these analyses point to MHC class II region,  
18 with *DRB1*, *DQA1*, and *DPA1* as the most likely candidate genes, but the extended LD across the region,  
19 combined with population differences in haplotype diversity and limitations related to imputation  
20 preclude further dissection of this complex signal.

21

## 22 **Pleiotropic associations of individual IgAN loci**

23

24 To describe the full spectrum of pleiotropic associations of individual risk variants, we cross-annotated all  
25 non-HLA signals against all genome-wide association studies listed in the NHGRI GWAS Catalogue  
26 (**Supplementary Tables 10-11**). We identified both concordant and opposed associations for multiple

2 autoimmune and inflammatory diseases, suggesting that these conditions may share pathogenic pathways  
3 with IgAN. Among the loci with the highest level of pleiotropy were *HORMAD2/LIF* and *ZMIZ1/PPIF*.  
4 Other loci with autoimmune pleiotropy were *CARD9*, *TNFSF8/15*, *REL*, *OVOL1/RELA*, *IRF4/DUSP22*,  
5 and *IRF8*. Additional novel loci, including *TNFRSF13B*, *PF4V1/CXCL1*, *LY86/RREB1*, and *ETS1*, showed  
6 concordant effects on blood levels of distinct immune cell types and immunoglobulins suggesting that these  
7 loci may act through stimulation of immune cell proliferation and immunoglobulin production. When we  
8 expanded this analysis to all suggestive loci, we found that 14 of the 47 suggestive loci were associated  
9 with the same autoimmune or blood immune cell traits as the genome-wide significant loci, prioritizing  
10 these 14 loci for future follow-up studies (**Figure 1c**).

11

## 12 **Shared genetic architecture with serum IgA levels and related traits**

13

14 To interrogate shared susceptibility between IgAN and other common diseases, we explored genome-wide  
15 genetic correlations with selected immune, infectious, and cardio-metabolic traits using bivariate LD score  
16 regression (**Figure 2, Supplementary Table 12**)<sup>10</sup>. We found negative genetic correlations with primary  
17 sclerosing cholangitis ( $r_g = -0.37$ ,  $P = 4.1 \times 10^{-3}$ ), inflammatory bowel disease ( $r_g = -0.16$ ,  $P = 9.9 \times 10^{-3}$ ), and  
18 Crohn's disease ( $r_g = -0.17$ ,  $P = 1.0 \times 10^{-2}$ ), and positive correlations with pneumonia ( $r_g = 0.26$ ,  $P = 9.0 \times 10^{-4}$ )  
19 and urinary tract infection ( $r_g = 0.25$ ,  $P = 2.1 \times 10^{-3}$ ). After excluding HLA, we also observed a positive genetic  
20 correlation with serum IgA levels ( $r_g = 0.31$ ,  $P = 2.1 \times 10^{-3}$ ), allergy ( $r_g = 0.18$ ,  $P = 5.2 \times 10^{-3}$ ), and tonsillectomy  
21 ( $r_g = 0.17$ ,  $P = 0.036$ ), a procedure performed for recurrent pharyngeal infections and also sometimes used to  
22 treat relapsing IgAN<sup>11</sup>. We next performed look-ups of all independent IgAN risk alleles against our latest  
23 GWAS for serum IgA levels (**Supplementary Table 13**). Of 25 non-HLA IgAN risk loci, 9 were nominally  
24 ( $P < 0.05$ ) associated with increased serum IgA levels, all with concordant effects. Conversely, of 31  
25 significant loci for IgA levels 12 were nominally associated with the risk of IgAN, also with concordant  
26 effects. The intersection include four highly significant loci in both GWAS: *TNFSF13* (GWAS for IgA

2 levels  $P_{IgA-level}=9.4 \times 10^{-8}$ ), *TNFSF8/15* ( $P_{IgA-level}=3.2 \times 10^{-10}$ ), *OVOLI/RELA* ( $P_{IgA-level}=2.6 \times 10^{-22}$ ), and  
3 *LIF/HORMAD2* ( $P_{IgA-level}=6.7 \times 10^{-17}$ ). At the same time, the allelic effects at the HLA locus were either  
4 opposed or not associated with serum IgA levels, consistent with our genetic correlation analyses in which  
5 positive genetic correlation with IgA levels is significant only after exclusion of the HLA region.

6

## 7 **Mouse phenotypes support the role of dysregulated IgA production in IgAN**

8

9 We tested the candidate gene set defined by our significant GWAS loci for overlap with human ortholog  
10 gene sets producing 27 phenotype categories when genetically manipulated in mice. We observed top-most  
11 significant enrichments in ‘*Immune system phenotype*’ ( $P=1.3 \times 10^{-12}$ ) and ‘*Hematopoietic system*  
12 *phenotype*’ ( $P=3.2 \times 10^{-9}$ ) (**Supplementary Table 14**). Within these categories, we observed significant  
13 enrichments in gene sets whose disruption in mice were associated with ‘*Abnormal IgA levels*’ ( $P=6.4 \times 10^{-$   
14  $6$ ) (**Extended Data Figure 6**), including *TNFSF13*, *TNFSF13B*, *ITGAM*, *RELA*, *REL*, *CD28*, and *LYN*  
15 genes. These observations corroborate our findings of overlapping GWAS loci between serum IgA levels  
16 and IgAN and further highlight the role of dysregulated IgA production in the disease pathogenesis.  
17 Moreover, this analysis strongly supports the named genes as causal at the corresponding loci and nominates  
18 appropriate animal models for experimental follow-up.

19

## 20 **Global pathway and tissue/cell type enrichment analyses**

21

22 We next used several unbiased strategies to explore biological pathway and tissue enrichments using  
23 genome-wide approaches. Pathway-enrichment analysis using MAGMA<sup>12</sup> revealed 24 enriched gene sets  
24 (**Extended Data Figure 7**). The most strongly enriched GO terms after excluding HLA region were  
25 ‘*Immune System Processes*’ (enrichment  $P=1.4 \times 10^{-9}$ ) and ‘*Immune Response*’ (enrichment  $P=2.6 \times 10^{-9}$ ).  
26 Examination of genome-wide significant non-HLA loci revealed significant enrichments in pathways

2 involved in innate and adaptive immunity, with the most significant enrichment in the ‘*Cytokine-Cytokine*  
3 *Receptor Interactions*’ (enrichment  $P=4.0 \times 10^{-11}$ ) (**Figure 3a**), suggesting several cytokine ligand-receptor  
4 interactions may drive disease pathogenesis.

5

6 To map the most likely causal tissues and cell types, we partitioned SNP-based heritability across the  
7 genome by tissue and cell-type-specific functional scores derived using the FUN-LDA method<sup>13</sup>. We found  
8 the most significant heritability enrichments in blood, immune, and gastrointestinal mucosa cells (**Figure**  
9 **3b, Supplementary Table 15**). The top enriched cell-types were *Primary neutrophils from peripheral*  
10 *blood* ( $P=5.9 \times 10^{-10}$ ), *PMA-I-stimulated primary T helper cells* ( $P=2.1 \times 10^{-9}$ ) and *Primary B cells from*  
11 *peripheral blood* ( $P=2.0 \times 10^{-8}$ ). Analogous analysis performed using experimental mouse datasets pointed  
12 to *Small intestine inflammatory cells under basal conditions and after Salmonella infection* as the top tissue  
13 (**Extended Data Figure 7**). Additional independent analytical methods (DEPICT<sup>14</sup> and GARFIELD<sup>15</sup>)  
14 similarly prioritized extra-renal tissues as likely causal in IgAN, converging on hematopoietic, immune,  
15 and gastrointestinal tissues as the most likely tissues to harbor causal cell types (**Figures 3c and 3d,**  
16 **Supplementary Tables 16-17**).

17

## 18 **Transcription factor enrichment analysis**

19 We tested for potential intersection of GWAS signals with a comprehensive database of transcription factor  
20 (TF) ChIP-seq datasets using the Regulatory Element Locus Intersection (RELI) algorithm<sup>16</sup>. In the  
21 analysis of genome-wide significant and suggestive loci, we detected significant intersection with binding  
22 sites for up 32 TFs in 52 immune cell types, with the most significant enrichments for RELA (corrected  
23  $P=5.3 \times 10^{-13}$ ) and NFKB1 (corrected  $P=1.9 \times 10^{-12}$ , **Figure 4d, Supplementary Table 18**). Nearly half of  
24 these transcription factors interact with Epstein-Barr-virus super-enhancers, which control B cell  
25 proliferation and have previously been found to intersect multiple autoimmune disease loci<sup>16,17</sup>. Moreover,  
26 some of the prioritized TFs, such RUNX<sup>18</sup> and SMAD<sup>19</sup> family, are well known to regulate IgA levels, and

2 RUNX3, RUNX2, and RELA loci are significantly associated with IgA levels (*see accompanying*  
3 *manuscript*), further suggesting perturbations in IgA homeostasis as a primary pathogenetic factor IgAN.

4

## 5 **Protein-protein interactions and ligand-receptor pairs**

6

7 We next tested whether candidate genes within our significant loci encode proteins that are likely to have  
8 physical interactions. Using a refined database of high-confidence PPIs, we constructed a network with 76  
9 candidate proteins defined by GWAS using InWeb\_IM<sup>20</sup> and GeneMANIA<sup>21</sup>. The final network composed  
10 of a total of 53 nodes and 63 edges exhibited an excess of direct physical interactions compared to null  
11 expectation ( $P < 1.0 \times 10^{-16}$ , **Figure 4c**). Gene set enrichment analyses of individual modules in this network  
12 (**Supplementary Table 19**) identified strong enrichments in stress and defense responses (module 1),  
13 chemokine signaling pathways (module 4), immune responses (module 5), cytokine-mediated signaling  
14 (module 6), and regulation of NF- $\kappa$ B signaling (module 7). Consistent with the observed enrichments in  
15 chemokine and cytokine pathways and global cytokine-receptor interactions, we identified enrichment in  
16 soluble ligand-receptor pairs, attributable to 16 ligand-receptor pairs spanning 12 independent significant  
17 or suggestive loci (enrichment  $P = 0.01$ , **Supplementary Table 20**). This included APRIL and its receptor  
18 TACI encoded by two independent genome-wide significant loci (*TNFSF13* and *TNFRSF13B*,  
19 respectively), both implicated in IgA homeostasis. Several IL6-related cytokine-receptor pairs were also  
20 identified (IL6-IL6ST, LIF-LIFR/IL6ST, OSM-OSMR/LIFR/IL6ST), with *OSM/LIF* being encoded by a  
21 single genome-wide significant locus, and related receptors being encoded by two independent suggestive  
22 loci, *OSMR/LIFR* and *IL6ST*. Notably, APRIL is known to alter the glycosylation of IgA<sup>22</sup>, IL6, LIF and  
23 OSM are involved in mucosal immunity, and IL6 and LIF leads to enhanced production of galactose-  
24 deficient IgA1<sup>23-25</sup>. These ligand-receptor pairs nominate candidate genes within corresponding loci, and  
25 delineate potentially targetable pathogenetic pathways in IgAN.

26

## 2 **Functional annotations of individual GWAS loci**

3

4 We intersected our association loci with tissue and cell-type specific functional scores and assessed their  
5 co-localization with expression quantitative trait loci (eQTL) in primary immune cells, whole blood, and  
6 other tissues (see **Methods**)<sup>26</sup>. We also performed cross-annotation with blood proteome and metabolome  
7 data. The majority of top signals mapped to non-coding regions, with the exception of two risk loci  
8 (rs4077515 *CARD9* p.(Ser12Ile), and rs3803800 *TNFSF13* p.(Asn96Ser)). The *CARD9* risk allele  
9 (rs4077515-T), a nonsynonymous S12N substitution in exon 2 of *CARD9*, is associated with increased  
10 blood transcript level of *CARD9* and a significant splice QTL in GTEx (**Extended Data Figure 8**). The  
11 protective allele is associated with a truncation of the functional CARD domain, while the risk allele is  
12 associated with higher levels of the intact, active isoform, affecting both expression and splicing of *CARD9*.

13

14 For top signals mapping to non-coding regions, we found 79 significant *cis*-eQTL effects with 17 IgAN co-  
15 localizations at 20 independent non-HLA risk loci (**Supplementary Tables 21-22 and Figure 4a**). Twelve  
16 loci had 27 significant *cis*-eQTL effects across 13 primary immune cell types, and 17 of the 27 *cis*-eQTLs  
17 co-localized with IgAN with PP4>0.5 (**Supplementary Table 21**). In GTEx, we further found 19 *cis*-eQTL  
18 effects for eight IgAN loci across the 28 available tissues and cell types. As an example, two loci  
19 (*ITGAM/ITGAX* and *IRF4/DUSP22*) mapped specifically to monocytes, an understudied cell type in IgAN.  
20 The top signals at these loci intersect monocyte-specific functional elements by FUN-LDA, and co-localize  
21 with monocyte-specific eQTLs, with the risk alleles associated with up- and down regulation of *ITGAX* and  
22 *IRF4/DUSP22*, respectively. As another example of cell type specificity, the *ZMIZ1/PPIF* locus co-  
23 localized with eQTL in NK cells, with the risk allele associated with lower expression of *ZMIZ1*, which  
24 encodes an inhibitor of JAK/STAT signaling and is also involved in TGF- $\beta$  signaling and intestinal  
25 inflammation<sup>27,28</sup>. In whole blood, notable eQTL co-localizations included the *FCRL3* risk locus, where the  
26 risk allele was associated with reduced transcript levels of *FCRL3* and *FCRL5*, and with lower levels of

2 circulating FCRL3 protein (**Supplementary Table 23**). As FCRL3 is a specific receptor for secretory  
3 IgA<sup>29,30</sup>, we prioritized *FCRL3* is the most likely causal gene at this locus.

4

5 Three independent IgAN risk loci with colocalizing cis-eQTLs also exhibited trans-eQTL effects,  
6 suggesting that these loci induce a more global transcriptional perturbation in blood cells (**Supplementary**  
7 **Table 24**). For example, the *CARD9* locus was associated with 12 trans-eQTL effects, nine of which involve  
8 genes in the ‘*Type I interferon signaling pathway*’ (enrichment  $P=9.5 \times 10^{-18}$ ). The *TNFSF8/15* locus was  
9 associated with eight trans-eQTL effects with three representing ‘*Cytokines involved in lymphocyte*  
10 *differentiation*’ (enrichment  $P=4.3 \times 10^{-3}$ ). Interestingly, *ITGAX* locus had only one trans-eQTL association,  
11 lowering mRNA level of *IGHG4*, encoded by an independent IgAN risk locus on chr.14.

12

13 Finally, other loci were associated with perturbations in blood proteome or metabolome. The *PF4VI* locus  
14 colocalized with *PF4VI* cis-eQTL and exhibited multiple pQTL associations with blood protein levels  
15 (**Supplementary Table 23**), including 4 cis and 40 trans-pQTLs proteins. These proteins were most  
16 enriched in the GO process of ‘*Positive Regulation of Neutrophil Chemotaxis*’ (enrichment  $P=1.3 \times 10^{-3}$ ),  
17 providing further evidence as *PF4VI* as the most likely causal gene for IgAN<sup>31</sup>. Similarly, the *CFH* locus,  
18 where a protective allele is known to tag a common deletion of the *CFHR1* and *CFHR3* genes<sup>2</sup>, was  
19 associated with reduced expression of *CFHR1* and *CFHR3* in the liver and other tissues including the kidney  
20 (**Supplementary Tables 25-26**). This allele was also associated with reduced levels of circulating FHR1  
21 (encoded by *CFHR1*) and higher levels of Factor H in blood (**Supplementary Table 23**). Moreover, this  
22 locus exhibited a widespread proteomic and metabolomic signature in blood, with 64 additional trans-pQTL  
23 associations including seven proteins involved in the ‘*Regulation of complement cascade*’ (enrichment  
24  $P=2.1 \times 10^{-10}$ , **Figure 4b**), and altered blood levels of multiple inflammation-related metabolites  
25 (**Supplementary Table 27**)<sup>32,33</sup>.

26

## 2 **Integrative prioritization of biological candidate genes**

3

4 To systematically prioritize the 311 candidate genes encoded within the 25 significant non-MHC risk loci,  
5 we scored for convergence of *in silico* annotation methods by assigning one point for each of the following  
6 criteria: 1) genes most proximal to the top SNP at the locus; 2) genes with a non-synonymous coding  
7 variant tagged ( $r^2 \geq 0.8$ ) by the top SNP; 3) genes with a 3D chromatin interaction predicted by the Activity-  
8 by-Contact (ABC) model<sup>34</sup> or 4) GeneHancer<sup>35</sup>, with enhancers that are intersected by variants tagged  
9 ( $r^2 \geq 0.8$ ) by the top SNP or contained within a 95% credible set for the locus; 5) e-genes controlled by at  
10 least one eQTL (any GTEx tissue) tagged by the top SNP; 6) e-genes co-localizing with the risk locus in  
11 peripheral blood or 7) primary immune cells at  $PP4 > 0.5$ ; 8) p-genes encoding blood proteins controlled by  
12 at least one cis-pQTL tagged by the top SNP; 9) genes prioritized by PPI network connectivity analysis at  
13  $P < 0.05$ ; 10) genes with shared mouse knockout phenotypes; 11) genes within shared MAGMA pathways;  
14 12) genes prioritized by DEPICT, and 13) genes prioritized by manual review of the literature as related to  
15 IgAN, IgA production, or mucosal immunity. Using this approach, we prioritized 27 ‘biological candidate  
16 genes’, 20 (74%) of which were also most proximal genes to the top SNP (**Figure 5**).

17

## 18 **Prioritization of plausible drug targets**

19

20 To facilitate drug repurposing and to prioritize new targets with GWAS support, we evaluated whether any  
21 of the 311 genes contained within significant loci encoded a protein or directly interacted with a protein  
22 that was a pharmacologically active drug target either approved or in development for any human disease.  
23 In total 13 GWAS loci (52%) encoded 17 proteins that were already targeted by existing drugs, and 11 loci  
24 (44%) encoded 14 proteins with a direct PPI drug target (**Supplementary Table 28 and Figure 6**). Among  
25 the top 27 high priority ‘biological candidates’ defined by our scoring system, 11 (40%) were targeted  
26 directly or indirectly by the existing drugs. This included the following drug categories: (1) new inhibitors

2 of the alternative complement pathway, such as APL-2, AMY-101, and several others<sup>36</sup> that are currently  
3 in clinical trials for C3 glomerulopathies and age-related macular degeneration; (2) drugs that block the  
4 activation of B cells by inhibiting APRIL or TACI interactions such as Atacicept and related drugs that are  
5 already in clinical trials for IgAN; (3) drugs that inhibit T cell activation by targeting ligands of the T-cell  
6 stimulatory CD28 protein, such as Belatacept (approved for kidney allograft rejection) or Abatacept  
7 (approved for rheumatoid arthritis); (4) drugs that inhibit IL8 (ABX-IL8) or IL8 receptor (Clotrimazole);  
8 and (5) drugs that inhibit NF- $\kappa$ B pathway, such as Bardoxolone that is already in clinical trials for  
9 glomerular disorders. We also note that some of our top prioritized causal genes with expression increased  
10 by the risk alleles, such as CARD9, ITGAX, PF4V1, CFHR1, or FCAR do not yet have effective drug  
11 inhibitors. Other loci encode secreted proteins that appear protective, such as FCRL3 and TNFSF4,  
12 suggesting that targeting their upregulation may present a rational therapeutic strategy. Our data  
13 additionally implies that activation of transcriptional programs controlled by ZMIZ1 and IRF4, but reduced  
14 activation of NF- $\kappa$ B, may also convey a protective effect.

15

## 16 **Genome-wide polygenic risk and clinical correlations**

17

18 Based on GWAS summary statistics after excluding ImmunoChip cohorts, we designed and optimized a  
19 genome-wide polygenic risk score (GPS) for IgAN. The best-performing GPS was based on LDpred  
20 method and assumed 1% causal variants genome-wide. When tested in the independent GCKD Study<sup>37,38</sup>,  
21 the GPS explained approximately 7.3% of disease risk ( $P=3.1 \times 10^{-12}$ , C-statistic 0.65, 95%CI: 0.61-0.68).  
22 We hypothesized that higher polygenic risk captured by the GPS was associated with disease severity and  
23 faster progression among cases diagnosed with IgAN. To test this hypothesis, we performed a  
24 comprehensive analysis of clinical disease features in association with the GPS (**Supplementary Table**  
25 **29**). Consistent with previous observations for the 15-SNP GRS<sup>4</sup>, the GPS was inversely associated with  
26 the age at diagnosis, with individuals in the top 20% tail of the GPS distribution diagnosed 2.2 years sooner

2 (95%CI: 1.3-3.1,  $P=2.5 \times 10^{-6}$ ) compared to the rest of the cohort. The GPS was also significantly associated  
3 with faster progression to kidney failure among 2,879 IgAN cases with long-term follow-up data (HR=1.17  
4 per standard deviation, 95%CI 1.09-1.24,  $P=3.3 \times 10^{-6}$ ). For example, individuals in the top 20% tail of the  
5 GPS distribution had 34% increased risk of kidney failure (HR=1.34, 95%CI=1.15-1.56,  $P=2.0 \times 10^{-4}$ ), while  
6 individuals in the top 10% tail had 48% increased risk of kidney failure (HR=1.48, 95%CI 1.22-1.79,  
7  $P=6.6 \times 10^{-5}$ ) compared to the rest of the cohort (**Figure 7a**).

8  
9 To explore additional clinical associations of the GPS, we performed meta-phenome-wide association study  
10 (meta-PheWAS) including a total of 590,515 participants with GWAS data linked to electronic health  
11 records, combining UK Biobank and Electronic Health Records and Genomics-III (eMERGE-III)  
12 consortium datasets (**Figure 7b**). We detected a significant positive correlation of the GPS with hematuria,  
13 the most common manifestation of IgAN (OR per standard deviation=1.06,  $P=7.3 \times 10^{-21}$ ). Other notable  
14 associations included a protective association with celiac disease (OR<sub>SD</sub>=0.60,  $P=4.2 \times 10^{-148}$ ) and several  
15 risk associations, including with rheumatoid arthritis (OR<sub>SD</sub>=1.15,  $P=1.1 \times 10^{-39}$ ), hypothyroidism  
16 (OR<sub>SD</sub>=1.05,  $P=2.0 \times 10^{-15}$ ), epistaxis or throat hemorrhage (OR<sub>SD</sub>=1.09,  $P=2.6 \times 10^{-9}$ ), and asthma  
17 (OR<sub>SD</sub>=1.02,  $P=1.5 \times 10^{-6}$ ). The above associations remained significant after removing the HLA region from  
18 the GPS (**Figure 7c, Supplemental Table 30**). Notably, the directions of effect were generally consistent  
19 with our genome-wide genetic correlation analyses of IgAN with related traits, providing an independent  
20 validation of the shared polygenic architecture for these traits.

21

22

## 23 **DISCUSSION:**

24

25 Our GWAS of 10,146 cases and 28,751 controls provided support for a highly polygenic architecture of  
26 IgAN with estimated SNP-based heritability of ~23%. We identified 16 novel susceptibility loci for

2 IgAN, bringing the total number of known risk loci to over 30 and explaining ~11% of the variance in  
3 disease risk. The higher polygenic risk was associated with earlier disease onset and greater lifetime risk  
4 of kidney failure, suggesting that polygenic background is predictive of a more aggressive disease. Future  
5 studies are needed to test if our polygenic stratification is useful in the diagnosis, clinical risk assessment,  
6 or prediction of treatment responsiveness.

7

8 Our results reinforce the hypothesis that the genetic regulation of IgA production represents the key  
9 pathogenic pathway in IgAN. Significant risk loci were enriched in human orthologs of mouse genes that,  
10 when genetically modified, cause abnormal IgA levels. We also observed positive genetic correlation  
11 between IgAN and serum IgA levels. Moreover, 21 of 25 independent genome-wide significant non-HLA  
12 risk loci for IgAN appear to have concordant effect on serum IgA levels, and four of these 21 are also  
13 genome-wide significant in a GWAS for serum IgA levels, including loci encoding TNFSF8, APRIL,  
14 LIF, and RELA.

15

16 We observed positive genetic correlations with IgA levels, infections, and tonsillectomy indicating a  
17 genetic link between IgA system, common infections, and IgAN. The association with tonsillectomy is  
18 especially intriguing, because IgAN is often triggered by pharyngitis, and tonsillectomy has been  
19 employed as a treatment for IgAN<sup>11</sup>. In contrast, the observed negative genetic correlations with  
20 inflammatory bowel disease may be due to genetically increased production of secretory IgA that has  
21 known homeostatic anti-inflammatory and immunosuppressive effects at the level of the gut mucosa<sup>39</sup>.  
22 Moreover, our analyses of partitioned heritability clearly support extra-renal tissues as the most likely  
23 culprit, prioritizing cells of the immune and hematopoietic systems, and intestinal mucosal tissue, and this  
24 extra-renal mapping of causal tissues is fully consistent with the well-established clinical observation that  
25 IgAN commonly recurs in kidney allografts after transplantation<sup>40</sup>. In addition, our cell-type specific

2 analyses point to the role of neutrophils, monocytes, and NK cells in IgAN. These cell types have not  
3 been generally considered as relevant to the IgAN pathogenesis based on prior evidence.

4

5 Our GWAS loci encoded proteins that were more likely to interact physically despite being encoded by  
6 distant genomic regions. This included several ligand-receptor pairs that are amenable to therapeutic  
7 targeting. IgAN currently lacks effective targeted therapies, and recent pharmaceutical database analyses  
8 indicate that drug targets with genetic support are more likely to advance in the development pipeline<sup>41</sup>.  
9 Based on our results, we prioritized several candidate genes whose products are targeted by drugs that are  
10 presently approved or in clinical development for another condition, and which could be repurposed for  
11 IgAN. Mechanistic studies are still needed to confirm our candidate target genes prioritized by our *in*  
12 *silico* annotations.

13

14 Our study has several limitations. In the meta-analysis, we pooled data across multiple heterogenous  
15 cohorts recruited across diverse clinical settings, ancestries, and nationalities. Nevertheless, we used  
16 stringent biopsy-based diagnostic criteria, standardized covariate definitions, genetic matching by  
17 platform and ancestry, and uniform statistical analysis for each cohort. It is worth pointing out that our  
18 dataset is dominated by European (66%) and East Asian (34%) ancestry cases, therefore our results may  
19 not be generalizable to other patient populations. Notably, IgAN is less frequent among individuals of  
20 African ancestry, including African Americans, suggesting that protective genetic effects may exist, but  
21 further studies are needed to address this hypothesis. We were also not able to evaluate the contribution of  
22 rare variants in this study, and sequencing studies are still needed to evaluate relative contributions of rare  
23 and common variants to the overall disease risk.

24

25

26

27

28

29

## 2 ONLINE METHODS

3

### 4 Study Cohorts, Genotyping, Genotype Quality Control and Imputation

5 The recruitment of patients, genotyping, imputation, and detailed quality control analyses are described by  
6 cohort in the **Supplemental Notes**.

7

### 8 Association Analyses and Meta-analyses

9 We conducted genome-wide association analysis in each of the 17 cohorts using imputed genotype dosage  
10 data under a logistic regression additive model with adjustment for cohort-specific significant PCs in  
11 PLINK v1.9<sup>42</sup>. Subsequently, a fixed effects inverse-variance-weighted meta-analysis was performed to  
12 combine results from all cohorts using METAL version 2011-03-25<sup>43</sup>. Genome-wide distributions of p-  
13 values were examined visually using quantile-quantile (QQ) plots for each individual cohort as well as for  
14 the combined analysis. We also estimated the genomic inflation factors for each cohort. The final meta-  
15 analysis QQ plot showed no departure from the expected distribution of p-values, and the genomic inflation  
16 factor ( $\lambda$ ) was estimated at 1.04 (**Extended Data Figure S1**). In addition, logistic regression association  
17 analysis assuming a dominant or a recessive genetic model was performed in PLINK using expected  
18 genotype counts, with no evidence of genomic inflation ( $\lambda=1.03$  for dominant and 0.94 for recessive model)  
19 (**Extended Data Figure S3**). IgAN is more common in males; thus, we also performed sex-specific  
20 analyses, including chromosome X. Genome-wide logistic regression analyses were conducted separately  
21 in males and females within each cohort and subsequently meta-analyzed using METAL. A total of 21,236  
22 males and 17,661 females were used in the meta-analysis with overall genomic inflation factors of 1.01 for  
23 males and 0.99 for females. A total of 1,990,322 high quality imputed chromosome X markers ( $r^2>0.8$  and  
24 MAF $>0.01$ ) were analyzed separately by sex, encoding genotypes as (0, 2) in males and (0, 1, 2) in females.  
25 Significant PCs for each cohort were included as covariates in each model. We defined a locus as genome-

2 wide significant if at least one SNP in the locus had  $p\text{-value} \leq 5.0\text{E-}08$  and it was successfully typed or  
3 imputed in  $\geq 50\%$  of analyzed cohorts. Signals with a  $p\text{-value} \leq 1.0\text{E-}05$  were considered as suggestive.  
4 Summary of meta-analysis results are provided in **Table 1 and Supplementary Tables S2-4**.

5

## 6 **Conditional Analyses**

7 To detect multiple independent associations at individual genomic loci, we conducted a stepwise  
8 conditional analysis using a multi-SNP-based conditional and joint association analysis (COJO)<sup>44</sup>. This  
9 method approximates the variance-covariance matrix between association statistics with the LD  
10 information from an external reference panel and independent SNPs are selected in a stepwise manner  
11 using the GCTA tool version 1.92.0beta<sup>44,45</sup>. Using our overall meta-analysis summary statistics, we  
12 conducted the GCTA conditional analysis with a threshold of  $P \leq 5.0\text{E-}08$  and the LD reference composed  
13 of all European and East Asian cohorts from the 1000 Genomes Project Phase 3. Subsequent conditional  
14 analyses were performed for makers with a conditioned  $P \leq 5.0\text{E-}08$  until no residual genome-wide  
15 significant associations were observed (**Supplementary Table S5**).

16

## 17 **HLA Imputation and Statistical Analysis**

18 We used the SNP2HLA software to impute classical HLA alleles<sup>46</sup>. The Type 1 Diabetes Genetics  
19 Consortium (T1DGC) reference panel of 5,225 Europeans and 8,961 markers was used as a reference set  
20 for our European-ancestry cohorts<sup>46</sup>, and the Pan-Asian reference panel of 530 individuals and 8,245  
21 markers was used for our East-Asian cohorts<sup>47</sup>. Only common and high-quality markers ( $\text{MAF} > 0.01$ ,  
22  $R^2 > 0.8$ ) were used for association analysis. We analyzed each variant using a logistic regression model,  
23 assuming additive dosage effects and controlling for significant PCs of ancestry. For testing multi-allelic  
24 loci, we used the following logistic regression model:

$$\log(\text{odds}) = \beta_0 + \sum_{j=1}^{m-1} \beta_j X_{j,i} + \sum_{k=1}^n \beta_k P_{k,i}$$

2  
3 where  $m$  indicates a total number of alleles at a specific multi-allelic locus,  $j$  indicates a specific allele  
4 being tested, and  $X_{j,i}$  is the imputed dosage for allele  $j$  for individual  $i$ ;  $\beta_0$  represents the intercept and  $\beta_j$   
5 represents the additive effect of an allele  $j$ ;  $P_{k,i}$  denotes the value for  $k^{\text{th}}$  PC of individual  $i$ ,  $n$  is the total  
6 number of significant PCs in the dataset;  $\beta_k$  is the effect size of principal component  $k$ . We compared log-  
7 likelihoods of two nested models: the full model containing the test locus (fitted model) and relevant  
8 covariates with the reduced model (null model) without the test locus, but with the same set of covariates.  
9 The deviance ( $D$ ) was defined as  $-2 \times \log$  likelihood ratio, which follows a  $\chi^2$  distribution with  $m-1$   
10 degrees of freedom, from which we calculated p-values:

$$D = -2 \ln \frac{L_0}{L_1}, D \sim \chi^2 (m - 1)$$

11  
12 where  $D$  is the log-likelihood test value (deviance),  $L_0$  is likelihood of the null model and  $L_1$  the likelihood  
13 of the fitted model. To identify statistically independent effects, we first tested all bi-allelic and multi-  
14 allelic variants under the logistic regression model, as described above, and ranked them based on the p-  
15 value of the log likelihood test. Next, in a forward stepwise approach, we included in the logistic  
16 regression model the most statistically significant variant as a covariate, analyzed all remaining variants  
17 and ranked them based on the new p-value of the respective log likelihood test. We repeated the same  
18 steps until no variant or no HLA gene had  $P \leq 5.0E-08$ .

19

## 20 **HLA Peptide Sequence Analysis**

21 HLA amino-acid polymorphisms have multiple possible residues at each peptide position. To test the  
22 effects of individual amino acid substitution sites we applied a conditional haplotype analysis using fully  
23 phased haplotypes across the HLA region. We tested each single amino acid position by first identifying

2 the  $m$  possible amino-acid residues occurring at that position and then using  $m-1$  degrees of freedom test  
3 to derive p-values with a single amino-acid residue arbitrary selected as a reference. For conditioning on  
4 individual amino-acid sites, we used the following procedure: by adding a new amino-acid position to the  
5 model, a total of  $k$  additional unique haplotypes were generated and tested over the null model (without a  
6 new amino-acid position) using the likelihood ratio test with  $k$  degrees of freedom. If the new position  
7 was independently significant, we further updated the null model to include all unique haplotypes created  
8 by all amino-acid residues at both positions to identify another independent position. The procedure was  
9 repeated until no statistically significant (conditioned  $P \leq 5.0E-08$ ) position was observed.

10

## 11 **Heritability and Genetic Correlations**

12 SNP-based heritability was estimated using LD score regression (LDSC software)<sup>10</sup> based on final meta-  
13 analysis summary statistics and LD scores estimated from 1000 Genomes phase 3 European and East  
14 Asian populations combined<sup>48</sup>. To assess the contribution of HLA region, we also estimated SNP  
15 heritability after excluding the entire MHC region (Chr.6: 28,000,000-33,000,000 bp). To investigate  
16 evidence for possible shared genetic effects between IgAN and other traits, we estimated genetic  
17 correlations using bivariate LD score regression<sup>10</sup>. For each phenotype, we used GWAS summary  
18 statistics from the largest GWAS available with a minimum coverage of 2 million SNPs. We excluded  
19 traits with estimated SNP-based heritability  $< 1\%$ . Genetic correlations were calculated with and without  
20 the HLA region. GWAS summary statistics for the relevant immune and cardio-metabolic traits were  
21 provided by corresponding consortia, or downloaded from the LD-hub or GWAS catalog; summary  
22 statistics for infection-related phenotypes were provided by 23andMe<sup>49</sup>.

23

## 24 **Pleiotropy Maps**

2 All GWAS loci were cross-annotated against the studies listed in the GWAS catalogue (last update: January  
3 31, 2019). For each locus, we selected all variants in strong LD ( $r^2 \geq 0.8$ ) with the top SNP. We then queried  
4 the GWAS catalogue for genome-wide significant ( $p < 5.0E-08$ ) associations of the selected SNPs with other  
5 diseases and traits. The results were then manually verified by reviewing the original publications to  
6 confirm the direction of allelic effects. In cases where there were multiple GWAS for the same trait, we  
7 selected the SNP associations based on the largest sample size/lowest P-value and highest LD with our top  
8 SNP(s). To evaluate the overlap of pleiotropic effects between significant and suggestive IgAN loci, the  
9 traits associated with significant IgAN loci were queried themselves against GWAS catalogue for  
10 associations with any of the suggestive SNPs or their proxies. The results were represented as a shared  
11 susceptibility network map created in Cytoscape v3.7.0 software.

12

### 13 **Polygenic Risk Models**

14 To assess the cumulative effect of independent significant and suggestive loci, we performed a genetic  
15 risk score (GRS) analysis. We first created two new GRS models based on the combined meta-analysis  
16 results: the 30-SNP model which comprises 30 independent genome-wide significant SNPs (25 non-HLA  
17 plus 5 HLA SNPs), and the 77-SNP GRS model which includes the same 30 SNPs plus 47 additional  
18 independent SNPs representative of the suggestive loci ( $p < 1.0E-05$ ). Each GRS was defined as the sum of  
19 the number of risk alleles weighted by their effect sizes. We required imputation  $R^2 > 0.3$  for including a  
20 SNP in the GRS calculation; individuals typed with ImmunoChip were excluded. Each GRS was  
21 standardized using a Z-score transformation with the mean and standard deviation of the control  
22 distribution. We evaluated the performance of each GRS by estimating two goodness of fit measures,  
23 Nagelkerke's pseudo  $R^2$  and the area under the receiver operating characteristics curve (AUROC).  
24 Genome-wide polygenic score (GPS) was calculated using LDpred<sup>50</sup> and LD-pruning and p-value  
25 thresholding (P+T) methods, similar to the methods in recent studies<sup>51,52</sup>. We used the combined meta-

2 analysis including high quality imputed SNPs that overlapped across all cohorts (2,408,512 SNPs) but  
3 excluding ImmunoChip cohorts. In the LDpred method, the genetic architecture prior for variant effect  
4 sizes is a Gaussian distribution that has two parameters: heritability and a fraction of causal variants.  
5 Heritability was estimated using LD score regression, and the fraction of causal variants was used as a  
6 tuning parameter ( $\rho$ ) across the following range: 1, 0.3, 0.1, 0.03, 0.01, 0.003, 0.001, 3.0E-04, 1.0E-04,  
7 1.0E-05, 1.0E-06. Using LDpred, a genome-wide predictor was calculated for each value of  $\rho$  and the best  
8 performing score was selected. The P+T method prunes variants in LD and considers only variants with a  
9 p-value under a certain threshold. Using a range of different  $r^2$  (0.2, 0.4, 0.6, 0.8) and p-value thresholds  
10 (1, 0.3, 0.1, 0.03, 0.01, 0.003, 0.001, 3.0E-04, 1.0E-04, 3.0E-05, 1.0E-05, 1.0E-06, 1.0E-07, 5.0E-08,  
11 1.0E-08), we again selected the best performing model. The performance of 30-SNP GRS, 77-SNP GRS,  
12 and best GPS were compared to the previously published 15-SNP GRS<sup>4</sup>. We additionally tested these  
13 models in the German Chronic Kidney Disease (GCKD) study<sup>37,38</sup>, including 314 histologically  
14 confirmed IgAN cases versus 663 disease controls with a biopsy-diagnosed kidney disease of another cause  
15 (see **Supplemental Note**). The analyses were implemented in R v.3.5.2 software.

16

## 17 **Gene Set and Pathway Enrichment Analyses**

18 We defined each IgAN-associated genomic locus by first selecting all proxy SNPs in LD ( $r^2 \geq 0.5$ ) with the  
19 lead SNP, then extending the genomic region 250 kb upstream and downstream of the first and last proxy  
20 SNP based on genomic position. Each region was then annotated using BiomaRt package, which retrieves  
21 Ensembl human gene annotations. Gene sets were created for all genome-wide significant and suggestive  
22 loci but excluding the HLA region. For gene set enrichment analysis, we used established gene sets from  
23 the Molecular Signatures Database (MSigDB), including GO, KEGG, BioCarta, REACTOME, chemical  
24 and genetic perturbations (CGP), and transcription factor targets (TFT). Statistical significance for  
25 enrichment was set at FDR q-value < 0.05. We additionally applied a genome-wide gene set enrichment

2 testing approach (excluding the HLA region) using Multi-marker Analysis of GenoMic Annotation  
3 (MAGMA) method with default parameters<sup>12</sup>. We also used DEPICT v1 release 194<sup>14</sup> to perform  
4 pathway/gene set enrichment and tissue/cell-type analyses. For this analysis, we first used PLINK to  
5 identify independently associated SNPs setting  $P < 5.0E-05$  and  $r^2 < 0.05$  in a physical window of 500 kb.  
6 We then used DEPICT to prioritize genes, identified reconstituted gene sets that are enriched in genes  
7 from associated regions, and identified tissue and cell-type annotations in which genes from associated  
8 regions are highly expressed. Specifically, for each tissue, the DEPICT method performs a *t*-test  
9 comparing the tissue-specific expression of IgAN-associated genes versus all other genes. Next, for each  
10 tissue, empirical enrichment p-values are computed by repeatedly sampling random sets of loci from the  
11 entire genome to estimate the null distribution for the enrichment statistic as previously described<sup>53,54</sup>.

12

### 13 **Prioritization of Causal Tissues and Cell Types**

14 We estimated heritability enrichment of SNPs from GWAS summary statistics for functional categories in  
15 tissue and cell-type specific regulatory elements using stratified LD score regression. This method  
16 regresses the chi-squared statistics of SNPs from summary statistics on their LD scores<sup>10</sup>, and partitions  
17 SNP heritability by functional annotation<sup>55</sup>. We used summary statistics obtained from the meta-analysis  
18 of all case-control cohorts excluding the MHC region and excluding the cohorts typed with Immunochip.  
19 Heritability enrichment was defined as the proportion of SNP heritability in a specific category, divided  
20 by the proportion of SNPs that belong to that category. We first calculated heritability enrichment for a  
21 baseline model of 96 non-cell-type-specific functional categories. Along with the 1000 Genomes  
22 reference panel, we used this baseline model as a control to assess heritability enrichment in cell-type  
23 specific functional annotations. We used functional annotations from the ENCODE and Roadmap  
24 Epigenomics Consortium<sup>56</sup>, as well as mouse immune cell-specific functional categories from the  
25 Immunological Gene Project (ImmGen)<sup>57</sup>. Using the same method, we evaluated tissue- and cell-specific

2 heritability enrichments based on the FUN-LDA functional scoring system<sup>13</sup>. As an alternative method,  
3 we also used GARFIELD v2<sup>15</sup> to assess enrichment within the ENCODE and Roadmap-derived  
4 regulatory regions.

5

## 6 **Analysis of Relevant Phenotypes in Mice**

7 As a complementary approach to prioritize genes within each locus, we used the Mouse Genome  
8 Informatics (MGI) database to identify potential genes the disruption of which causes relevant phenotypes  
9 in mice<sup>58</sup>. All phenotypes in MGI are categorized based on the Mammalian Phenotype (MP) ontology and  
10 emerge as a result of different genetic models, including targeted knockout animals and chemically  
11 induced (ENU) and spontaneous mutations. MGI includes a total of 17,101 mouse genes with Human  
12 orthologs<sup>59</sup>. We defined a gene set of 62 genes with mouse orthologs across the 24 non-HLA risk loci for  
13 testing against MGI phenotypes to define significantly enriched categories.

14

## 15 **Functional Annotations of Individual Loci**

16 For the purpose of detailed functional annotation, we calculated 95% credible sets for each of the genome  
17 wide significant loci using CAVIAR software<sup>60</sup>. We added variants that were neither typed or imputed in  
18 our data, but in strong LD with the top SNP based on external reference ( $r^2 \geq 0.8$  in 1000G European and  
19 East Asian populations). These SNP sets were annotated using ANNOVAR to first define any coding  
20 variants and their predicted effects. Using FUN-LDA method, we next calculated the posterior probability  
21 of a functional effect for each of the selected variants across 127 tissues and cell types as described  
22 previously<sup>13</sup>. These SNPs were also interrogated against the following datasets: (1) eQTLs for 13 human  
23 immune cell types from the Database of Immune Cell eQTLs (DICE) project<sup>61</sup>; (2) blood eQTLs from the  
24 eQTLGen Consortium<sup>26</sup> (31,684 individuals), (3) GTEx tissue eQTLs<sup>62</sup> (4) GTEx splicing QTLs (sQTLs);  
25 (5) kidney eQTLs (glomerular and tubular) from the Kidney eQTLs Atlas<sup>63</sup>, (6) blood mQTLs from KORA

2 (N=2,820) and TwinsUK (N=7,824) studies<sup>64,65</sup>, and (7) blood pQTLs from three recent well-powered  
3 studies<sup>66-68</sup>. We additionally performed co-localization analysis between IgAN and eQTL summary  
4 statistics for each GWAS locus using COLOC software<sup>69</sup>. We considered PP4>0.5 as supportive of a shared  
5 causal variant.

6

## 7 **Protein-Protein Interactions**

8 Protein-protein interactions across genome wide significant and suggestive genes have been predicted  
9 using InWeb\_InBioMap (InWeb\_IM)<sup>70</sup>. InWeb\_IM is a curated and computationally derived protein-  
10 protein network of 420,000 protein-protein interactions that has 2.8 times more interactions than other  
11 comparable resources. We used high confidence PPIs from InWeb\_IM using the recommended cut off  
12 confidence score  $\geq 0.1$ . All annotated genes within the 76 significant and suggestive IgAN loci were used  
13 to probe the PPI database; the final network contained a total of 53 nodes connected by 63 edges.  
14 Enrichment P-value was computed using a hypergeometric test and corrected for multiple testing using  
15 the method of Benjamini and Hochberg. The network components were grouped into modules based on  
16 their pathway categorization; the GLay community clustering algorithm was implemented for module  
17 detection and modules were visualized in GeNets<sup>71</sup>. Subsequently, the Clustering with Overlapping  
18 Neighborhood Expansion (ClusterONE) algorithm<sup>72</sup> implemented in Cytoscape<sup>73</sup> was used to extract  
19 protein clusters using the default parameters and the InWeb\_IM confidence score as edge weights.  
20 Functional and pathway enrichments within the PPI networks were identified using STRING<sup>74</sup> based on  
21 Gene Ontology (GO), KEGG, and Reactome databases. In addition, we used ToppGene Suite<sup>75</sup> to  
22 calculate protein-protein interaction enrichment p-values. A Bonferroni-corrected  $P < 0.05$  was used as  
23 enrichment significance cut-off.

24

## 25 **Transcription Factor-DNA Binding Interactions**

2 To identify TF binding sites enriched across IgAN risk loci, we used the RELI (Regulatory Element  
3 Locus Intersection) algorithm<sup>16</sup>. RELI uses a set of genetic variants as input, expands the set using LD  
4 blocks ( $r^2 > 0.8$ ) and calculates the statistical intersection of the resulting loci with ChiP-seq datasets by  
5 counting the number of loci with one or more variants intersecting the TF ChiP-seq peaks. The LD blocks  
6 were calculated using 1000 Genomes Project East Asian and European populations combined. The null  
7 distribution was generated using 2,000 random repeats of the procedure and was used to calculate z-scores  
8 and empirical p-values for the observed intersection. The final reported p-values are Bonferroni-corrected  
9 ( $P_c$ ) for the 1,544 TF datasets tested, as previously published<sup>16</sup>. As input, we used a set of 28 independent  
10 genome-wide significant loci ( $P \leq 5.0E-08$ ) and a set of 76 loci including genome-wide significant and  
11 suggestive variants ( $P < 1.0E-05$ ). A  $P_c < 1.0E-04$  was used as a significance cut-off for each set.

12

### 13 **Ligand-Receptor Pairs**

14 To identify the number of potential ligand-receptor pairs significantly associated with IgAN, we queried  
15 The Database of Ligand-Receptor Partners (DLRP)<sup>76</sup>. This database includes cytokines, chemokines, and  
16 growth, angiogenesis, and developmental factors, and contains 175 protein ligands, 131 protein receptors  
17 and 451 experimentally determined ligand-receptor pairings. To test for ligand-receptor enrichment in our  
18 dataset, we used a hypergeometric test for overlap between this dataset and the gene set defined by our  
19 significant and suggestive GWAS loci.

20

### 21 **Analysis of Drug Targets**

22 We obtained drug target genes and corresponding drug information from DrugBank<sup>77</sup>, the Therapeutic  
23 Targets Database (TTD)<sup>78</sup>, the Open Targets Platform<sup>79</sup>, and GlobalData combined with manual literature  
24 searches. To search for potential drug targets, we extracted all genes in direct PPIs with IgAN risk genes  
25 by using the In\_Web\_IM database. We selected drug target genes that had pharmacological activities and

2 human orthologues, and that were targeted by any of the drugs that are approved or currently in  
3 development (experimental or in clinical trials).

#### 4 5 **Overall Prioritization of Biological Candidate Genes**

6 Each of the positional candidate genes was scored adopting the following criteria and calculating the  
7 number of the satisfied criteria, including: (1) genes most proximal to the top SNP; (2) genes with coding  
8 variants in 95% credible sets and/or high LD ( $r^2 > 0.8$ ) with the index SNP; (3) genes with promoter  
9 chromatin interaction by Activity-by-Contact (ABC) model<sup>34</sup> or (4) GeneHancer<sup>35</sup> involving regions  
10 intersected by top SNP and its 95% credible sets/high LD proxies; (5) e-genes controlled by at least one  
11 eQTL (any tissue) tagged by the top SNP in any tissues (primary immune cells, whole blood, kidney,  
12 GTEx); (6) e-genes co-localized with the risk locus in blood or (7) primary immune cells with  $PP4 > 0.5$ ;  
13 (8) p-genes controlled by at least one blood pQTL tagged by the top SNP; (9) genes prioritized by PPI  
14 network connectivity analysis at  $P < 0.05$ ; (10) genes that when knocked out in mice produce at least two  
15 phenotype labels: ‘immune system’, ‘haematopoietic system’, or ‘cellular phenotype’; (11) genes  
16 prioritized by MAGMA, (12) DEPICT with gene-based  $p < 0.05$ , or (13) manual review of the literature as  
17 related to IgAN, IgA production, or mucosal immunity.

#### 18 19 **Genotype–phenotype Correlations**

20 Our polygenic risk models (15-SNP, 30-SNP, 77-SNP, and GPS) were tested for clinical correlations in  
21 the subset of cases with available clinical data. We fitted each standardized risk score in a regression  
22 model to predict clinical disease features at the time of diagnosis, including age at biopsy, estimated  
23 glomerular filtration rate (GFR), proteinuria, microhematuria, hypertension (HTN), and gross hematuria.  
24 The GFR was estimated using the CKD-EPI formula in adults<sup>80</sup> and Schwartz formula in pediatric  
25 cases<sup>81</sup>. The estimated GFR was normalized with a natural logarithm transformation; proteinuria was

2 normalized with a  $\ln(P24+1)$  transformation; microhematuria was assessed with urinary dipstick and  
3 defined as positive if 1+ or greater; gross hematuria was defined by the reported history of red urine  
4 before a diagnostic biopsy; hypertension was defined as systolic blood pressure  $\geq 140$  mmHg and/or  
5 diastolic blood pressure  $\geq 90$  mmHg, or anti-hypertensive medication use. For the analysis of longitudinal  
6 outcomes, we performed survival analysis with the primary outcome of kidney failure defined as  
7  $eGFR < 15 \text{ ml/min/1.73 m}^2$  or initiation of kidney replacement therapy (dialysis or kidney transplantation).  
8 All analyses were adjusted for age, gender, site, and race/ethnicity. The analyses were implemented in R  
9 v3.5.2.

10

## 11 **Phenome-wide association studies (PheWAS)**

12 We performed meta-analysis of PheWAS results (meta-PheWAS) across two large biobank-based  
13 datasets: Electronic Medical Records and Genomics-III (eMERGE-III) and the UK Biobank (UKBB).  
14 The eMERGE-III network provides access to EHR information linked to GWAS data for 102,138  
15 individuals; detailed quality control analyses of genetic data have been described previously<sup>82-84</sup>. Briefly,  
16 GWAS datasets were imputed using the latest multiethnic Haplotype Reference Consortium (HRC) panel  
17 using Michigan Imputation Server<sup>85</sup>. The imputation was performed in 81 batches across the 12  
18 contributing medical centers participating in eMERGE-I, II, and III. For post-imputation analyses, we  
19 included only markers with minor allele frequency (MAF)  $\geq 0.01$  and  $R^2 \geq 0.8$  in  $\geq 75\%$  of batches. A  
20 total of 7,529,684 variants were retained for the GPS analysis. For principal component analysis (PCA),  
21 we used FlashPCA<sup>86</sup> on a set of 48,509 common (MAF $>0.01$ ) and independent variants (pruned in  
22 PLINK with `--indep-pairwise 500 50 0.05` command). The UKBB is a large prospective population-based  
23 cohort that enrolled individuals ages 40-69 for the purpose of genetic studies<sup>87</sup>. This cohort is comprised  
24 of 488,377 individuals recruited since 2006, genotyped with high-density SNP arrays, and linked to  
25 electronic health record data. All individuals underwent genome-wide genotyping with UK Biobank

2 Axiom array from Affymetrix and UK BiLEVE Axiom arrays (~825,000 markers). Genotype imputation  
3 was carried out using a 1000 Genomes reference panel with IMPUTE4 software<sup>88-90</sup>. We then applied QC  
4 filters similar to eMERGE-III, retaining 9,233,643 common ( $MAF \geq 0.01$ ) variants imputed with high  
5 confidence ( $R^2 \geq 0.8$ ). For principal component analysis by FlashPCA<sup>86</sup>, we used a set of 35,226 variants  
6 that were common ( $MAF > 0.01$ ) and pruned using the following command in PLINK --indep-pairwise  
7 500 50 0.05. In order to test the GPS for associations phenome-wide across both eMERGE and UKBB  
8 datasets, we first harmonized the coded diagnoses data by converting all available ICD-10-CM codes to  
9 ICD-9-CM system. This approach was motivated by the fact that the great majority of data for eMERGE-  
10 III participants is already coded using ICD-9 system, and ICD-10 codes are more granular, thus a reverse  
11 conversion leads to mapping errors. After the conversion, the 102,138 genotyped eMERGE participants  
12 had a total of 20,783 ICD-9 codes that were then mapped to 1,817 distinct phecodes (disease-specific  
13 groupings of ICD codes). The 488,377 UKBB participants had a total of 10,221 ICD-9 codes that that  
14 were mapped to 1,817 phecodes. Phenome-wide associations were performed using the PheWAS R  
15 package<sup>91</sup>. The package uses pre-defined “control” groups for each phecode. The case definition requires  
16 a minimum of two ICD-9 codes from the “case” grouping of each phecode. In total, all 1,817 phecodes  
17 were tested using logistic regression with each phecode case-control status as an outcome and the  
18 polygenic score IgA nephropathy adjusted for age, sex, study site or batch, and 3 principal components of  
19 ancestry as a predictor. We then performed meta-PheWAS across both datasets combined using metagen  
20 with fixed effect in PheWAS R library<sup>91</sup>. To establish significant disease associations in PheWAS, we set  
21 the Bonferroni-corrected statistical significance threshold at  $2.75E^{-05}$  ( $0.05/1,817$ ) correcting for 1,817  
22 independent phecodes tested.

23  
24  
25  
26

## 2 Web Resources

3

4 **COLOC:** <https://cran.r-project.org/web/packages/coloc/>

5 **DICE:** <https://dice-database.org/>

6 **EAGLE:** <https://data.broadinstitute.org/alkesgroup/Eagle/>

7 **eQTLGen:** <https://www.eqtlgen.org/>

8 **DEPICT:** <https://data.broadinstitute.org/mpg/depict/index.html>

9 **DLRP:** <https://dip.doe-mbi.ucla.edu/dip/DLRP.cgi>

10 **DrugBank:** <https://www.drugbank.ca>

11 **FUN-LDA:** <http://www.columbia.edu/~ii2135/funlda.html>

12 **GARFIELD:** <https://www.ebi.ac.uk/birney-srv/GARFIELD/>

13 **GCTA-COJO:** <https://cnsgenomics.com/software/gcta/#COJO>

14 **GeNets:** <http://apps.broadinstitute.org/genets>

15 **GlobalData:** <https://www.globaldata.com/industries-we-cover/pharmaceutical>

16 **GSEA:** <http://software.broadinstitute.org/gsea/msigdb/>

17 **GTEx:** <https://gtexportal.org/home/>

18 **GWAS catalog:** <https://www.ebi.ac.uk/gwas>

19 **InWeb:** [http://apps.broadinstitute.org/genets#InWeb\\_InBiomap](http://apps.broadinstitute.org/genets#InWeb_InBiomap)

20 **KING:** <http://people.virginia.edu/~wc9c/KING/>

21 **Kidney eQTL Atlas:** <http://susztaklab.com/eqtl>

22 **LD hub:** <http://ldsc.broadinstitute.org/ldhub>

23 **LDpred:** <https://github.com/bvilhjal/ldpred>

24 **LDSC:** <https://github.com/bulik/ldsc>

25 **Metabolomics GWAS Server:** <http://metabolomics.helmholtz-muenchen.de/gwas/>

26 **METAL:** <http://csg.sph.umich.edu/abecasis/Metal/>

27 **MGI:** <http://www.informatics.jax.org>

28 **MINIMAC3:** <http://genome.sph.umich.edu/wiki/Minimac3>

29 **MIP:** <https://imputationserver.sph.umich.edu>

30 **NEPTUNE eQTL Browser:** <http://nephqtl.org/>

31 **Open Targets:** <https://www.targetvalidation.org>

32 **PheWAS:** <https://github.com/PheWAS/PheWAS>

33 **PheWeb:** <http://pheweb.sph.umich.edu/SAIGE-UKB/>

34 **PLINK:** <https://www.cog-genomics.org/plink/1.9/>

35 **RELI:** <https://github.com/WeirauchLab/RELI>

36 **SNP2HLA:** [http://software.broadinstitute.org/mpg/snp2hla/snp2hla\\_manual.html](http://software.broadinstitute.org/mpg/snp2hla/snp2hla_manual.html)

37 **STRING:** <https://string-db.org>

38 **ToppGene:** <https://toppgene.cchmc.org>

39 **TTD:** <https://db.idrblab.org/ttd>

40

## 2 Acknowledgments

3 We are grateful to all study participants across multiple nephrology centers worldwide for their  
4 contributions to this manuscript. This work was supported by the following institutions, grants, and  
5 funding agencies: Columbia University, Columbia Glomerular Center, IGA Nephropathy Foundation of  
6 America, and National Institute for Diabetes and Digestive Kidney Diseases (NIDDK) grants R01-  
7 DK105124 (KK, JN, BAJ) and RC2-DK116690 (KK), R01-DK082753 (AGG, JN, KK, FS, BAJ, RJW).  
8 Additional support was provided by R01LM013061 (KK), U01HG008680 (KK), U01AI152960 (KK),  
9 R01-DK078244 (JN, BAJ), R01-AI149431 (JN, BAJ), National Science Foundation of China (82022010,  
10 ZX; 82070733, ZH); Beijing Natural Science Foundation (Z190023,ZX); DFG Project-ID 192904750 –  
11 CRC 992 Medical Epigenetics (AK, PS), DFG Project-ID 431984000 – CRC 1453 (AK), and the EQUIP  
12 Program for Medical Scientists, Faculty of Medicine, University of Freiburg (PS). JBH, SP, KL, and WM  
13 were supported by R01HG010730, U01AI130830, R01NS099068, R01AI024717, R01AR073228,  
14 U01AI150748, R01AI148276, P01AI150585 and the US Department of Veterans Affairs Merit Award  
15 I01-BX001834 to JBH. DPG was supported by the St Peter’s Trust for Kidney, Bladder and Prostate  
16 Research. The UK cohort data was generated as a result of a grant from Kidney Research UK and the  
17 Medical Research Council (JB and DG). The German STOP-IgAN study was supported by the Deutsche  
18 Forschungsgemeinschaft (DFG, German Research Foundation) – CRU 5011 – Project-ID 445703531  
19 (J.F., T.R.). The GCKD (German Chronic Kidney Disease) study was funded by grants from the German  
20 Ministry of Education and Research (BMBF, No. 01ER0804) and the KfH Foundation for Preventive  
21 Medicine. Unregistered grants to support the study were provided by Bayer, Fresenius Medical Care, and  
22 Amgen. Genotyping was supported by Bayer Pharma AG. The recruitment of Polish cases with IgA  
23 nephropathy was sponsored by the Polish Kidney Genetics Network (POLYGENES), a collaborative  
24 effort between Columbia University and Poznań University of Medical Sciences, Poland. The recruitment  
25 of Czech patients with IgA nephropathy was supported by the research project of General University  
26 Hospital in Prague (RVO-VFN64165). The recruitment of Russian cohort was supported by the

2 Government Assignment of the Russian Ministry of Health, Assignment #200080056 of the Veltishev  
3 Research and Clinical Institute for Pediatrics of the Pirogov Russian National Research Medical  
4 University. The recruitment of the Swedish cohort was supported by the Nephrology and Rheumatology  
5 Departments at Karolinska University Hospital and Karolinska Institutet, Stockholm, and grants from the  
6 Swedish Society of Medicine. We thank the Immunopathology Working Group of the Italian Society of  
7 Nephrology (ISN) for inviting their member sites to contribute to this study. We are also grateful to the  
8 Pediatric Nephrology Research Consortium (PNRC) for hosting and co-sponsoring the GIGA-kids Study  
9 that recruited pediatric patients with IgAN across PNRC sites. We are additionally grateful to Corinna  
10 Bowers for coordinating GIGA-kids and PMRC-based recruitment, Dr. Joann Narus for coordinating  
11 recruitment at the University of Utah, Dr. Ewa Elenberg and Dr. Shweta Shah for helping enroll subjects  
12 from Texas Children's Hospital and Adisak Suwanichkul for sample shipping. The recruitment of the  
13 Korean cohort was supported by the Seoul National University Hospital Human Biobank, a member of  
14 the National Biobank of Korea, financed by the Ministry of Health and Welfare, Republic of Korea. We  
15 are grateful to the International Multiple Sclerosis Genetics Consortium (IMSGC), the Inflammatory  
16 Bowel Disease Genetics Consortium (IBDGC), the International Myositis Consortium (IMC), the  
17 Feinstein Institute for Medical Research, the French Biological Resource Center for MS Genetics  
18 (REFGENSEP), Genethon, and INSERM for contributing ImmunoChip controls for the purpose of this  
19 study. We would also like to thank the Population Architecture Using Genomics and Epidemiology  
20 (PAGE) consortium, funded by the National Human Genome Research Institute (NHGRI) with co-  
21 funding from the NIMHD, for providing population controls genotyped with MEGA chip for this study.  
22 The funding sources were not involved in the study design, collection, analysis, and interpretation of data,  
23 writing of the report, or in the decision to submit the paper for publication.

24

25

## 2 **Authors Contributions**

3 K. Kiryluk and A.G. Gharavi conceived the study, provided overall supervision of the project, and made  
4 the decision to publish the findings; E. Sanchez-Rodriguez and N. Mladkova performed quality control,  
5 imputation, and GWAS association analyses for GWAS and ImmunoChip discovery cohorts. E. Sanchez-  
6 Rodriguez performed final statistical analyses, meta-analyses, fine-mapping studies, functional  
7 annotations, and drug target analyses; F. Zanon and A. Khan designed polygenic risk scores, tested for  
8 genome-wide genetic correlations, and performed clinical correlation analyses; L. Liu performed analysis  
9 of the HLA locus; J.B. Harley, M. Weirauch, S. Eswar, S. Parameswaran, and L. Kottyan performed  
10 RELI transcription factor analysis; I. Ionita-Laza consulted on the statistics and functional annotation of  
11 GWAS loci; K. Kiryluk, A.G. Gharavi, J. Novak provided biological interpretation of GWAS loci; X.  
12 Zhou and H. Zhang coordinated recruitment, genotyping, and analysis of the Beijing cohorts; F. Scolari  
13 coordinated recruitment activities across the Italian network of clinical recruitment sites; Y. Caliskan  
14 coordinated recruitment activities across the Turkish recruitment sites; H. Trimarchi coordinated  
15 recruitment across the Argentinian network of clinical recruitment sites; V. Tesar and D. Maixnerova  
16 coordinated the recruitment of the Czech cohort; J. Xie and N. Chan coordinated recruitment, genotyping,  
17 and analysis of the Shanghai cohorts; H. Suzuki recruited and clinically characterized the Japanese  
18 Discovery cohort. H. Lee, J. Park, B.L. Cho, Y.S. Kim, and D.K. Kim recruited and clinically characterized  
19 the Korean Discovery cohort. J. Floege and T. Rauen contributed DNA and clinical data for the German  
20 STOP-IgAN cohort; K. Kiryluk, R. Nelson, R. Wyatt, and B. Smoyer led the GIGA-kids study in  
21 collaboration with the Pediatric Nephrology Research Consortium (PNRC); M. Marasa, O. Balderes, J.Y.  
22 Zhang coordinated recruitment at Columbia University and managed clinical data and DNA samples; J.  
23 Barratt, and D. P. Gale provided summary statistics for the UK GWAS cohort; R. P. Lifton contributed  
24 previously published European and Beijing GWAS cohorts; A. B. Ekici, K-U. Eckardt, and A. Köttgen  
25 contributed German cases to the ImmunoChip discovery cohorts, and P. Schlosser performed polygenic  
26 risk score validation studies in the GCKD cohort; D. van Heel, B. Cotsapas, C. Wijmenga, A. Franke, V.

2 Annese, and P.K. Gregersen contributed Illumina ImmunoChip idat files for population controls; R. Loss  
3 and E. Kenny contributed Illumina MEGA chip idat files for population controls. All other co-authors  
4 contributed to the recruitment and clinical characterization of IgAN patients and controls recruited from  
5 their respective clinical centers. All authors have read and approved the final version of manuscript.

6

## 7 **Declaration of Interests**

8 B.A.J. and J.N. are co-founders, co-owners of, and consultants for Reliant Glycosciences, LLC and are  
9 co-inventors on US patent application 14/318,082 (assigned to UAB Research Foundation). A.G. has  
10 served on an advisory board for Novartis, Travere and Natera and receives research grant funding from  
11 the Renal Research Institute and Natera. K.K. has served on an advisory board for Goldfinch Bio and  
12 Gilead. The other authors report no conflict of interest.

13

## 14 **Ethics Statement**

15 All subjects provided informed consent to participate in genetic studies, and the Institutional Review  
16 Board of Columbia University (protocol number IRB-AAAC7385) as well as local ethics review  
17 committees for each of the individual cohorts approved our study protocol.

18

## 19 **Data Availability**

20 Primary genotype data for previously published cohorts is available through dbGAP under accession  
21 number phs000431.v2.p1. Genotype data for newly added cohorts will be available through dbGAP upon  
22 publication (accession number pending). Our IRB determined that the use of this dataset is restricted to  
23 genetic studies of kidney disease. The PAGE consortium control genotype data is available on dbGAP  
24 under accession number phs000356.v2.p1. The Electronic Medical Records and Genomics-III (eMERGE-

- 2 III) imputed genotype and phenotype data are available through dbGAP, accession number:
- 3 phs001584.v2.p2. The UK Biobank genotype and phenotype data are available through the UK Biobank
- 4 web portal <https://www.ukbiobank.ac.uk/>. All data and summary statistics are also available from the
- 5 corresponding authors upon reasonable request.

2 **Table 1. New and known genome-wide significant loci based on meta-analysis.**

Chr	Location (bp <sup>a</sup> )	SNP <sup>b</sup>	Locus	Risk Allele	Freq. European Controls	Freq. Asian Controls	OR	P-value	Q test	I <sup>2</sup>	Supporting Cohorts	New
1	157,542,162	rs849815	<i>FCRL</i>	A	0.66	0.50	1.14	3.9E-09	0.57	0	GWAS + IC	Known
1	173,146,357	rs4916312	<i>TNFSF4</i>	A	0.35	0.07	1.14	5.0E-08	0.68	0	GWAS + IC	New
1	196,686,918	rs6677604	<i>CFH</i>	G	0.80	0.93	1.21	1.5E-17	0.97	0	GWAS + IC	Known
1	196,603,302	rs12029571	<i>CFH</i>	A	0.22	0.35	1.12	2.5E-06	0.88	0	GWAS	Known
2	61,092,678	rs842638	<i>REL</i>	T	0.44	0.15	1.17	9.6E-10	0.002	63.8	GWAS	New
2	204,584,759	rs3769684	<i>CD28</i>	T	0.95	0.49	1.19	5.1E-11	0.93	0	GWAS + IC	New
4	74,725,320	rs6828610	<i>PF4V1/CXCL8</i>	G	0.16	0.27	1.14	3.5E-08	0.98	0	GWAS	New
6	249,571	rs12201499	<i>IRF4/DUSP22</i>	C	0.12	0.28	1.18	3.1E-11	0.39	5.1	GWAS	Known
6	7,214,676	rs12530084	<i>LY86</i>	C	0.77	0.51	1.13	1.3E-09	0.49	0	GWAS	New
6	32,389,305	rs9268557	<i>HLA-DRA</i>	C	0.51	0.57	1.24	4.5E-47	0.00	72.4	GWAS + IC	Known
6	32,667,829	rs9275355	<i>HLA-DQB/DQA</i>	C	0.23	0.33	1.26	1.7E-34	0.21	24.8	GWAS	Known
6	32,599,999	rs9272105	<i>HLA-DQA</i>	A	0.60	0.45	1.25	1.2E-28	0.001	73.2	GWAS + IC	Known
6	32,681,631	rs9275596	<i>HLA-DQB/DQA</i>	T	0.66	0.81	1.33	3.2E-36	0.05	45.9	GWAS + IC	Known
6	33,074,288	rs3128927	<i>HLA-DPA/DPB</i>	C	0.73	0.83	1.22	1.5E-25	0.38	6.1	GWAS + IC	Known
6	167,445,139	rs2282859 <sup>c</sup>	<i>CCR6</i>	C	0.01	0.15	1.20	3.9E-08	0.06	43.8	GWAS + IC	New
8	6,808,722	rs2075836	<i>DEFA</i>	T	0.31	0.30	1.21	5.8E-11	0.90	0	GWAS	Known
8	56,852,496	rs75413466	<i>LYN</i>	A	0.02	0.06	1.40	1.4E-10	0.80	0	GWAS + IC	New
8	124,765,474	rs34354351	<i>ANXA3</i>	T	0.17	0.32	1.15	3.5E-08	0.84	0	GWAS	New
9	117,643,362	rs13300483	<i>TNFSF8/15</i>	T	0.24	0.31	1.13	1.3E-08	0.88	0	GWAS + IC	New
9	139,266,496	rs4077515	<i>CARD9</i>	T	0.41	0.29	1.14	2.6E-11	0.41	1.8	GWAS + IC	Known
10	65,363,048	rs57917667	<i>REEP3</i>	G	0.02	0.19	1.22	1.1E-08	0.91	0	GWAS	New
10	81,043,743	rs1108618	<i>ZMIZ1</i>	A	0.60	0.49	1.14	1.9E-10	0.43	0.3	GWAS + IC	New
11	65,555,524	rs10896045	<i>RELA</i>	A	0.30	0.48	1.18	4.7E-13	0.03	50.6	GWAS	New
11	128,487,069	rs7121743	<i>ETS1</i>	C	0.16	0.47	1.13	3.4E-08	0.79	0	GWAS	New
14	107,222,014	rs751081288	<i>IGH</i>	A	0.43	0.53	1.17	1.9E-08	0.74	0	GWAS	New
16	31,357,760	rs11150612	<i>ITGAM</i>	A	0.64	0.27	1.16	8.4E-14	0.10	36.8	GWAS+IC	Known
16	86,017,715	rs1879210	<i>IRF8</i>	T	0.64	0.86	1.14	9.9E-09	0.96	0	GWAS+IC	New
17	7,462,969	rs3803800	<i>TNFSF13</i>	A	0.21	0.32	1.15	1.2E-10	0.19	27.1	GWAS+IC	Known
17	16,851,450	rs57382045	<i>TNFRSF13B</i>	A	0.11	0.33	1.16	3.4E-09	0.84	0	GWAS	New
19	55,397,217	rs1865097	<i>FCAR</i>	A	0.30	0.38	1.12	7.7E-09	0.49	0	GWAS+IC	New
22	30,512,478	rs4823074	<i>HORMAD2/LIF</i>	G	0.54	0.67	1.16	7.8E-15	0.51	0	GWAS	Known

3  
4 Chr., chromosome; freq., frequency; Q test, p-value for Cochran's Q statistic; I<sup>2</sup>, heterogeneity index; IC, Immunochip

5 <sup>a</sup> Genome Reference Consortium Human Build 37 (hg19).

6 <sup>b</sup> Only independent SNPs in each locus are included.

7 <sup>c</sup> East Asian genome-wide significant specific signal.

8

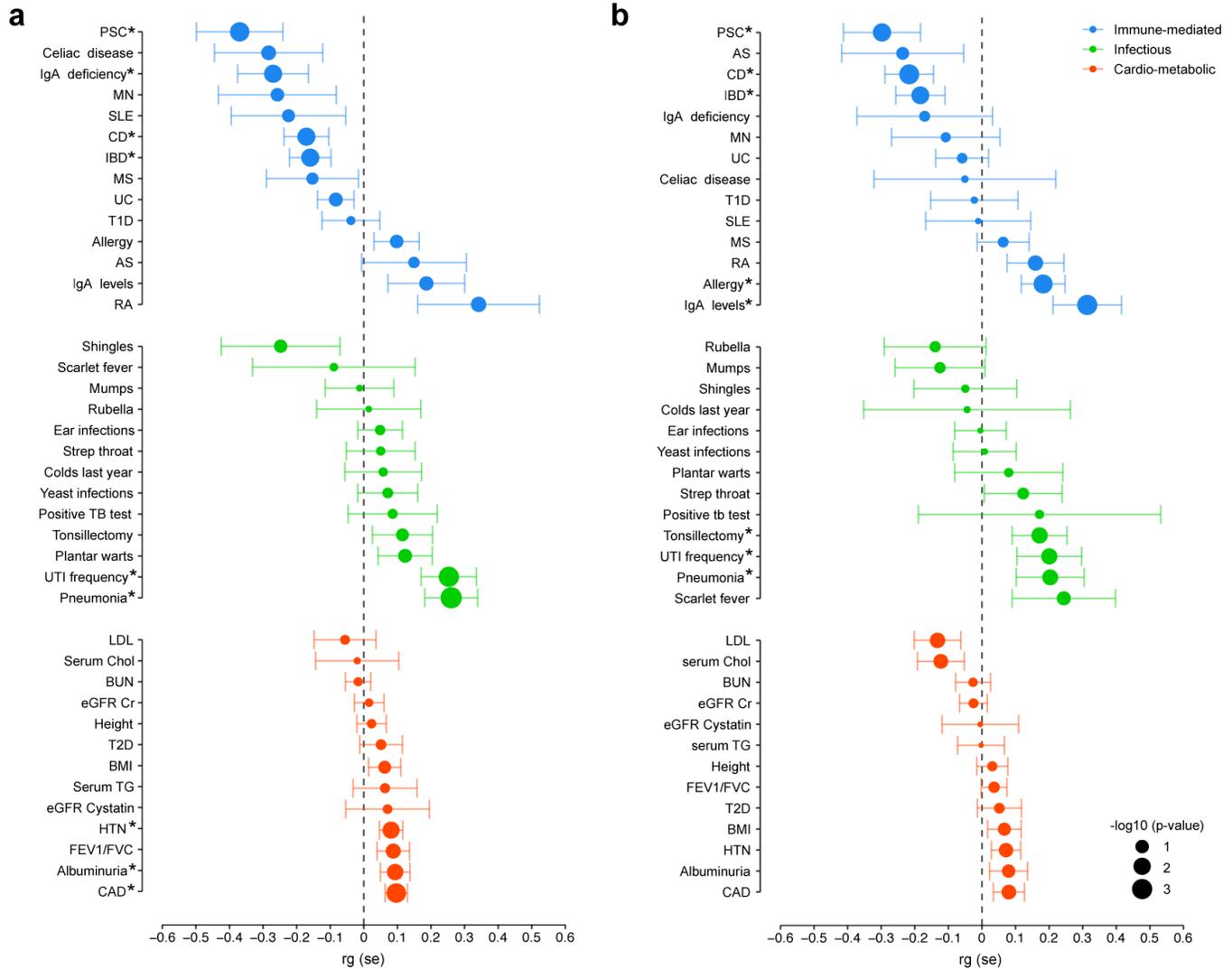
9

10



2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12

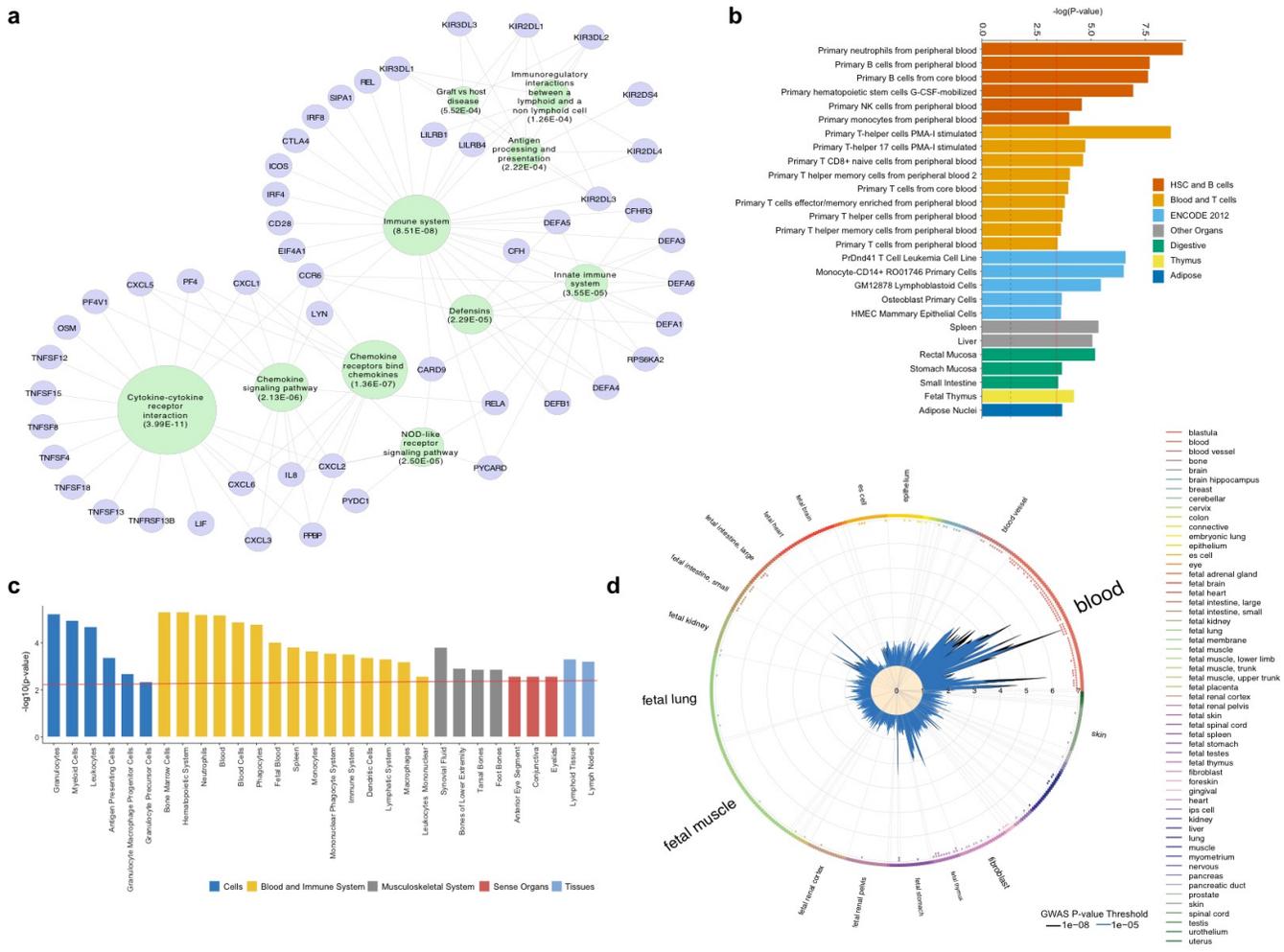
**Figure 2. Genome-wide genetic correlation analysis between IgA nephropathy and other complex traits (a) including HLA region and (b) excluding HLA region.** The traits are organized by immune-mediated (blue), infectious (green), and cardiometabolic (orange) categories and sorted based on the genetic correlation coefficient (rg); point estimates are provided with their 95% confidence intervals. PSC: primary sclerosing cholangitis, MN: membranous nephropathy, SLE: systemic lupus erythematosus, CD: Crohn's disease, IBD: inflammatory bowel disease, MS: multiple sclerosis, UC: ulcerative colitis, T1D: type 1 diabetes, AS: ankylosing spondylitis, RA: rheumatoid arthritis, Chol: total serum cholesterol levels, TG: total serum triglycerides levels, LDL: low-density lipoprotein levels, BUN: blood urea nitrogen, eGFR Cr: estimated glomerular filtration rate using serum creatinine levels, T2D: type 2 diabetes, BMI: body mass index, eGFR Cystatin: estimated glomerular filtration rate using serum cystatin levels, HTN: essential hypertension, FEV1/FVC: forced expiration volume at 1 second over forced vital capacity, CAD: coronary artery disease. \* phenotypic correlations at P<0.05.



13

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

2 **Figure 3. Global pathway, cell type, and tissue enrichment analyses:** (a) KEGG, REACTOME and BIOCARTA pathway  
 3 enrichment map based on the gene set defined by genome-wide significant IgA nephropathy loci after excluding HLA region; the  
 4 top 10 most significantly enriched pathways and their intersecting GWAS genes are shown; a node size reflects  $-\log_{10}$ -  
 5 transformed P-values of the multiple-testing adjusted hypergeometric enrichment test in GSEA. (b) Cell type-specific heritability  
 6 enrichment for functional annotations based on FUN-LDA scoring system for all ENCODE and Roadmap Epigenomics cell types  
 7 and tissues; the histogram shows the results of partitioned heritability by cell type-specific FUN-LDA functional annotations; only  
 8 significant results grouped by the tissue type are depicted; a solid red line represents the Bonferroni-corrected  $-\log_{10}$  of the p-  
 9 value threshold of significance ( $P=3.9 \times 10^{-4}$ ); a dotted black line represents the  $-\log_{10}$  of the nominal p-value of 0.05; the most  
 10 significant heritability enrichment were found in blood and immune cells and as well as gastrointestinal mucosal tissues. (c) Tissue  
 11 and cell-type enrichment analysis with DEPICT; only cells and tissues with a false discovery rate  $< 0.05$  are shown; Y-axis  
 12 represents the  $-\log_{10}$  of the p-value and x-axis shows the first level MeSH tissue and cell type annotations; the strongest  
 13 enrichment is observed for blood and immune cells. (d) Global GWAS enrichment in DNase I-hypersensitive sites (DHS) using  
 14 GARFIELD; radial lines show odds ratios at two GWAS p-values thresholds (T) for all DHS cells and tissues on the outer circle;  
 15 dots in the inner ring of the outer circle denote significance of GARFIELD enrichment at  $T < 1.0 \times 10^{-5}$  (outermost) and  $T < 1.0 \times 10^{-8}$   
 16 (innermost) after multiple-testing correction for the number of effective annotations. The dots colored with respect to the tissue  
 17 cell type tested (the font size of tissue labels reflects the number of cell types from that tissue, only tissues). Similar to FUN-LDA,  
 18 GWAS results are most enriched in DHS sites in blood and immune cells, and intestinal mucosal tissue.

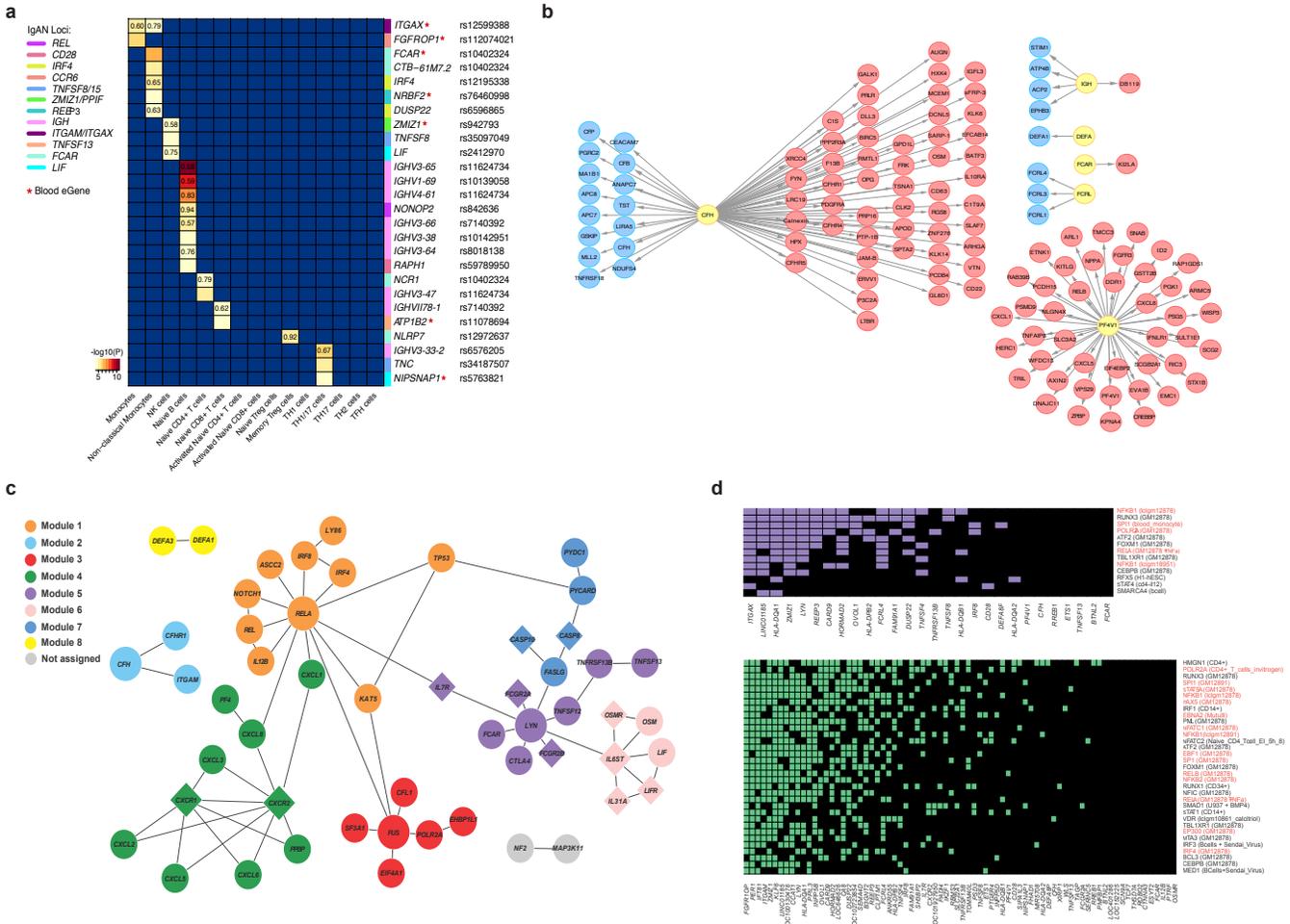


20  
 21  
 22  
 23  
 24

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

**Figure 4. Summary of candidate causal gene prioritizations across IgAN GWAS loci excluding HLA region: (a) cis-eQTL effects in primary immune cells: x-axis: 13 immune cell types of the Database of Immune Cell expression QTL epigenomics (DICE) project; posterior probability for a shared causal variant (PP4) is shown for the e-QTL effects that co-localize at  $PP4 > 0.50$ ; y-axis: significant eGene-eSNP pair with shared GWAS loci depicted by a color bar; \*eGenes with detectable blood cis-eQTLs. (b) pQTL effects in blood: IgAN risk alleles or their proxies (yellow nodes) with significant blood pQTLs depicted by blue (reduced protein levels) and red (increased protein levels) nodes; edge thickness corresponds to the LD between the lead pQTL SNP and the top IgAN GWAS SNP at a given locus. (c) Protein-protein interaction (PPI) network for candidate genes at GWAS loci based on InWeb\_IM: modules represent genes that are more connected to one another than they are to other genes; each module exhibits a functional enrichment network based on Gene Ontology (GO) biological processes: module 1 represents response to stress and defense response networks (orange), module 2 represents regulation of inflammatory response network (light blue), module 3 represents mRNA splicing via spliceosome network (red), module 4 represents chemokine-mediated signaling pathway network (green), module 5 represents immune response network (purple), module 6 represents cytokine-mediated signaling pathway network (pink), module 7 represents regulation of I-Kappa B kinase/NF-kappaB signaling and apoptotic signaling pathway networks (dark blue) and module 8 represents innate immune response in mucosa and antibacterial humoral response networks (yellow); the gray module has no functional enrichment. Overall, this network has more connectivity than expected by chance (permutation  $P < 2e-03$ ). (d) Intersection with transcription factor (TF) ChIP-seq peaks with the significant (top) and suggestive (bottom) IgAN risk loci; x-axis: IgAN risk loci; y-axis: top significant TFs ranked by the number of intersecting loci; a colored box at the intersection indicates that at a given locus has at least one IgAN-associated variant located within a ChIP-seq peak for the given TF; Datasets were considered significant if their RELI corrected p-values were less than  $1E-04$ ; TFs binding to EBNA2 super-enhancers colored in red; ChIP-seq dataset cell type indicated in parentheses; related cell lines for a given TF (e.g., GM12878 and GM12891) were merged for clarity.**



24  
25  
26  
27  
28  
29  
30

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

2  
3  
4  
5  
6  
7  
8  
10

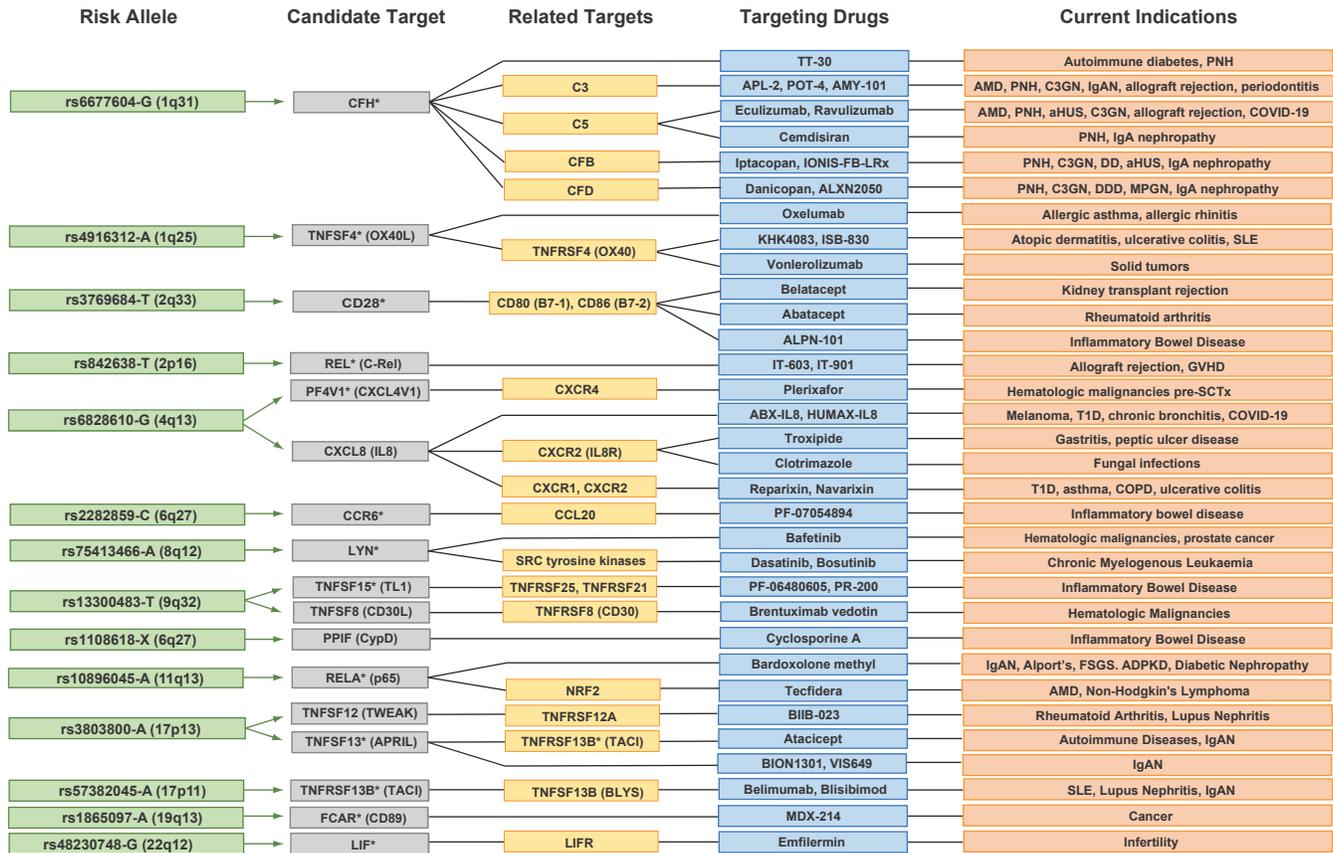
**Figure 5. Prioritization of candidate genes at non-HLA loci:** blue boxes indicate prioritization criteria based on genomic coordinates (the nearest gene to the index SNP, exonic variant in LD with the top SNP, or top signal intersecting chromatin interaction site with the gene promoter); red boxes indicate the presence of additional functional criteria (any GTEx eQTL effect, blood and immune cell eQTL co-localization, pQTL effects, PPI network connectivity, shared mouse KO phenotype, shared pathways by MAGMA, prioritized by DEPICT, and prioritized by manual PubMed review). The priority score represents a sum of the 13 scoring criteria depicted in blue and red. The genes with the maximum score at each locus (green boxes) were defined as 'biological candidate genes'. Additional annotation indicates drug target genes (yellow boxes). Only 56 of 311 positional candidate genes with a score >3 are depicted.

Locus	Top SNP	Candidate Gene	Nearest gene	Non-synonymous variant	ABC Promoter Interaction	GENEHANCER Interaction	eQTL in GTEx (any tissue)	Blood eQTL Co-localization	Immune Cells Co-localization	pQTL	PPI	Mouse KO phenotype	MAGMA pathways	DEPICT prioritized	PubMed Review	Priority Score	Drug target	PPI with drug target
1q31	rs6677604	CFH														10		
		CFHR1														5		
		CFHR3														4		
1q23	rs849815	FCRL3														7		
		FCRL4														5		
		FCRL5														4		
1q25	rs4916312	TNFSF4														8		
		TNFSF18														4		
2q33	rs3769684	CD28														7		
		CTLA4														5		
		RAPH1														4		
2p16	rs842638	REL													7			
4q13	rs6828610	PF4V1														9		
		CXCL1														8		
		CXCL8														5		
		CXCL6														4		
		CXCL5														4		
6p25	rs12201499	IRF4														7		
		DUSP22														6		
6p24	rs12530084	LY86														5		
		RREB1														4		
6q27	rs2282859	CCR6														6		
		FGFR10P														4		
8p23	rs2075836	DEFA1														4		
		DEFA4														4		
8q12	rs75413466	LYN													6			
8q24	rs34354351	ANXA13													4			
9q34	rs4077515	CARD9														10		
		INPP5E														5		
		SNAPC4														4		
9q32	rs13300483	TNFSF15														7		
		TNFSF8														6		
10q22	rs1108618	ZMIZ1														8		
		PPIF														6		
10q21	rs57917667	REEP3													5			
11q13	rs10896045	OVOL1														5		
		RELA														5		
		EHBP1L1														4		
11q24	rs7121743	ETS1													5			
14q32	rs751081288	IGH													6			
16p11	rs11150612	ITGAX														10		
		ITGAM														7		
		PYCARD														5		
		ITGAD														4		
		FUS														4		
16q24	rs1879210	IRF8													8			
17p13	rs3803800	TNFSF13														8		
		TNFSF12														6		
17p11	rs57382045	TNFRSF13B													7			
19q13	rs1865097	FCAR													6			
22q12	rs4823074	LIF														8		
		OSM														5		
		ASCC2														4		
		HORMAD2														4		

11

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

2 **Figure 6: Drug targets among candidate causal genes.** IgAN risk alleles (green), prioritized positional candidate genes (gray),  
 3 related genes in PPI (e.g. ligands/receptors) or same pathway (yellow), targeting drugs approved or currently in clinical trials  
 4 including agonists and antagonists (blue), and diseases targeted by these drugs (orange); high priority targets defined in Figure 5  
 5 are indicated by an asterisk; GWAS loci with candidate causal genes not targeted by the existing drugs are not depicted.  
 6

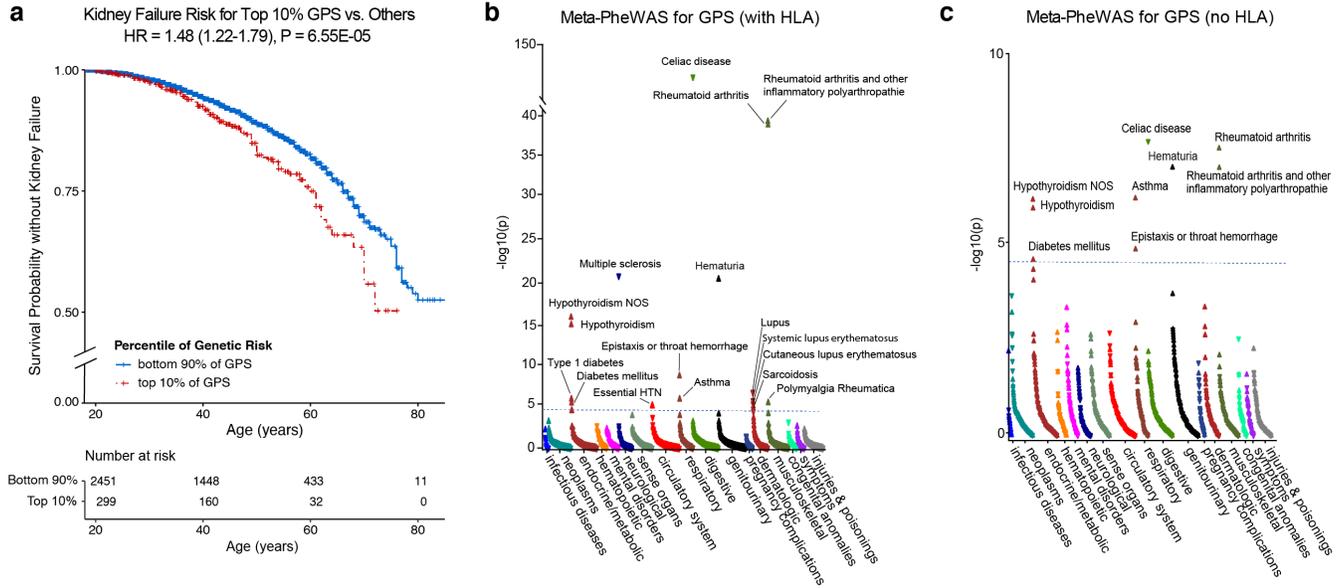


7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12

**Figure 7. Clinical associations of the genome-wide polygenic risk score (GPS) for IgA nephropathy.** (a) Survival analysis of a lifetime risk of kidney failure for IgAN cases in the top 90<sup>th</sup> percentile of the GPS distribution (N=2,879 cases with follow up data). The x-axis shows age starting from 18 years, the y-axis shows survival probability without kidney failure with the number of participants at risk at each age cut-off of 20, 40, 60, and 80 years depicted below; HR: hazard ratio (95% confidence interval) of kidney failure adjusted for sex, site, and ancestry. (b) Phenome-wide association study (PheWAS) for the GPS, including (c) the GPS without the HLA region, based on joint meta-analysis of eMERGE-III (N=102,138) and UKBB (N=488,377) datasets. The x-axis indicates electronic health record phenotypes (phecodes) grouped by system and sorted by significance. The y-axis indicates the level of statistical significance expressed as  $-\log_{10}(P\text{-value})$ . An upward triangle indicates a positive association (increased risk) and a downward triangle indicates a negative association (decreased risk) with increasing GPS. Dotted horizontal line represents the significance threshold after Bonferroni correction for the number of phenotypes; significant associations are labeled.



13  
14

2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46

## References:

- 1 Feehally, J. *et al.* HLA has strongest association with IgA nephropathy in genome-wide analysis. *Journal of the American Society of Nephrology : JASN* **21**, 1791-1797, doi:10.1681/ASN.2010010076 (2010).
- 2 Gharavi, A. G. *et al.* Genome-wide association study identifies susceptibility loci for IgA nephropathy. *Nat Genet* **43**, 321-327, doi:10.1038/ng.787 (2011).
- 3 Kiryluk, K. *et al.* Geographic differences in genetic susceptibility to IgA nephropathy: GWAS replication study and geospatial risk analysis. *PLoS Genet* **8**, e1002765, doi:10.1371/journal.pgen.1002765 (2012).
- 4 Kiryluk, K. *et al.* Discovery of new risk loci for IgA nephropathy implicates genes involved in immunity against intestinal pathogens. *Nature genetics* **46**, 1187-1196, doi:10.1038/ng.3118 (2014).
- 5 Yu, X. Q. *et al.* A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. *Nature genetics* **44**, 178-182, doi:10.1038/ng.1047 (2012).
- 6 Li, M. *et al.* Identification of new susceptibility loci for IgA nephropathy in Han Chinese. *Nature communications* **6**, 7270, doi:10.1038/ncomms8270 (2015).
- 7 Li, M. *et al.* Genome-Wide Meta-Analysis Identifies Three Novel Susceptibility Loci and Reveals Ethnic Heterogeneity of Genetic Susceptibility for IgA Nephropathy. *J Am Soc Nephrol* **31**, 2949-2963, doi:10.1681/ASN.2019080799 (2020).
- 8 Kiryluk, K., Novak, J. & Gharavi, A. G. Pathogenesis of immunoglobulin A nephropathy: recent insight from genetic studies. *Annu Rev Med* **64**, 339-356, doi:10.1146/annurev-med-041811-142014 (2013).
- 9 Kiryluk, K. & Novak, J. The genetics and immunobiology of IgA nephropathy. *J Clin Invest* **124**, 2325-2332, doi:10.1172/JCI74475 (2014).
- 10 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295, doi:10.1038/ng.3211 (2015).
- 11 Feriozzi, S. & Polci, R. The role of tonsillectomy in IgA nephropathy. *J Nephrol* **29**, 13-19, doi:10.1007/s40620-015-0247-4 (2016).
- 12 de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219, doi:10.1371/journal.pcbi.1004219 (2015).
- 13 Backenroth, D. *et al.* FUN-LDA: A Latent Dirichlet Allocation Model for Predicting Tissue-Specific Functional Effects of Noncoding Variation: Methods and Applications. *Am J Hum Genet* **102**, 920-942, doi:10.1016/j.ajhg.2018.03.026 (2018).
- 14 Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890, doi:10.1038/ncomms6890 (2015).
- 15 lotchkova, V. *et al.* GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. *Nat Genet* **51**, 343-353, doi:10.1038/s41588-018-0322-6 (2019).
- 16 Harley, J. B. *et al.* Transcription factors operate across disease loci, with EBNA2 implicated in autoimmunity. *Nat Genet* **50**, 699-707, doi:10.1038/s41588-018-0102-3 (2018).
- 17 Zhou, H. *et al.* Epstein-Barr virus oncoprotein super-enhancers control B cell growth. *Cell Host Microbe* **17**, 205-216, doi:10.1016/j.chom.2014.12.013 (2015).
- 18 Sugai, M., Watanabe, K., Nambu, Y., Hayashi, T. & Shimizu, A. Functions of Runx in IgA class switch recombination. *J Cell Biochem* **112**, 409-414, doi:10.1002/jcb.22971 (2011).

- 2 19 Malhotra, N. & Kang, J. SMAD regulatory networks construct a balanced immune system.  
3 *Immunology* **139**, 1-10, doi:10.1111/imm.12076 (2013).
- 4 20 Lage, K. *et al.* A human phenome-interactome network of protein complexes implicated in  
5 genetic disorders. *Nat Biotechnol* **25**, 309-316, doi:10.1038/nbt1295 (2007).
- 6 21 Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for  
7 gene prioritization and predicting gene function. *Nucleic Acids Res* **38**, W214-220,  
8 doi:10.1093/nar/gkq537 (2010).
- 9 22 Han, S. S. *et al.* The Role of TNF Superfamily Member 13 in the Progression of IgA Nephropathy. *J*  
10 *Am Soc Nephrol* **27**, 3430-3439, doi:10.1681/ASN.2015060677 (2016).
- 11 23 Yamada, K. *et al.* Leukemia Inhibitory Factor Signaling Enhances Production of Galactose-  
12 Deficient IgA1 in IgA Nephropathy. *Kidney Dis (Basel)* **6**, 168-180, doi:10.1159/000505748 (2020).
- 13 24 Suzuki, H. *et al.* Cytokines alter IgA1 O-glycosylation by dysregulating C1GalT1 and ST6GalNAc-II  
14 enzymes. *J Biol Chem* **289**, 5330-5339, doi:10.1074/jbc.M113.512277 (2014).
- 15 25 Yamada, K. *et al.* Inhibition of STAT3 Signaling Reduces IgA1 Autoantigen Production in IgA  
16 Nephropathy. *Kidney Int Rep* **2**, 1194-1207, doi:10.1016/j.ekir.2017.07.002 (2017).
- 17 26 Vosa, U. *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and  
18 polygenic scores that regulate blood gene expression. *Nat Genet* **53**, 1300-1310,  
19 doi:10.1038/s41588-021-00913-z (2021).
- 20 27 Owen, C. R., Yuan, L. & Basson, M. D. Smad3 knockout mice exhibit impaired intestinal mucosal  
21 healing. *Lab Invest* **88**, 1101-1109, doi:10.1038/labinvest.2008.77 (2008).
- 22 28 Yang, X. *et al.* Targeted disruption of SMAD3 results in impaired mucosal immunity and  
23 diminished T cell responsiveness to TGF-beta. *EMBO J* **18**, 1280-1291,  
24 doi:10.1093/emboj/18.5.1280 (1999).
- 25 29 Agarwal, S. *et al.* Human Fc Receptor-like 3 Inhibits Regulatory T Cell Function and Binds  
26 Secretory IgA. *Cell Rep* **30**, 1292-1299 e1293, doi:10.1016/j.celrep.2019.12.099 (2020).
- 27 30 Tolnay, M. Lymphocytes sense antibodies through human FCRL proteins: Emerging roles in  
28 mucosal immunity. *J Leukoc Biol*, doi:10.1002/JLB.4RU0221-102RR (2021).
- 29 31 Choobdar, S. *et al.* Assessment of network module identification across complex diseases. *Nat*  
30 *Methods* **16**, 843-852, doi:10.1038/s41592-019-0509-5 (2019).
- 31 32 Zhang, X. *et al.* An integrative investigation of the therapeutic mechanism of *Ainsliaea fragrans*  
32 Champ. in cervicitis using liquid chromatography tandem mass spectrometry based on a rat  
33 plasma metabolomics strategy. *J Pharm Biomed Anal* **156**, 221-231,  
34 doi:10.1016/j.jpba.2018.04.048 (2018).
- 35 33 Zhao, L. *et al.* Stachydrine ameliorates isoproterenol-induced cardiac hypertrophy and fibrosis by  
36 suppressing inflammation and oxidative stress through inhibiting NF-kappaB and JAK/STAT  
37 signaling pathways in rats. *Int Immunopharmacol* **48**, 102-109, doi:10.1016/j.intimp.2017.05.002  
38 (2017).
- 39 34 Fulco, C. P. *et al.* Activity-by-contact model of enhancer-promoter regulation from thousands of  
40 CRISPR perturbations. *Nat Genet* **51**, 1664-1669, doi:10.1038/s41588-019-0538-0 (2019).
- 41 35 Fishilevich, S. *et al.* GeneHancer: genome-wide integration of enhancers and target genes in  
42 GeneCards. *Database (Oxford)* **2017**, doi:10.1093/database/bax028 (2017).
- 43 36 Zipfel, P. F. *et al.* Complement Inhibitors in Clinical Trials for Glomerular Diseases. *Front Immunol*  
44 **10**, 2166, doi:10.3389/fimmu.2019.02166 (2019).
- 45 37 Titze, S. *et al.* Disease burden and risk profile in referred patients with moderate chronic kidney  
46 disease: composition of the German Chronic Kidney Disease (GCKD) cohort. *Nephrol Dial*  
47 *Transplant* **30**, 441-451, doi:10.1093/ndt/gfu294 (2015).

- 2 38 Wunnenburger, S. *et al.* Associations between genetic risk variants for kidney diseases and  
3 kidney disease etiology. *Sci Rep* **7**, 13944, doi:10.1038/s41598-017-13356-6 (2017).
- 4 39 Hansen, I. S., Baeten, D. L. P. & den Dunnen, J. The inflammatory function of human IgA. *Cell Mol*  
5 *Life Sci* **76**, 1041-1055, doi:10.1007/s00018-018-2976-8 (2019).
- 6 40 Infante, B. *et al.* Recurrence of immunoglobulin A nephropathy after kidney transplantation: a  
7 narrative review of the incidence, risk factors, pathophysiology and management of  
8 immunosuppressive therapy. *Clin Kidney J* **13**, 758-767, doi:10.1093/ckj/sfaa060 (2020).
- 9 41 Nelson, M. R. *et al.* The support of human genetic evidence for approved drug indications. *Nat*  
10 *Genet* **47**, 856-860, doi:10.1038/ng.3314 (2015).
- 11 42 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets.  
12 *Gigascience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).
- 13 43 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide  
14 association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).
- 15 44 Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies  
16 additional variants influencing complex traits. *Nat Genet* **44**, 369-375, S361-363,  
17 doi:10.1038/ng.2213 (2012).
- 18 45 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait  
19 analysis. *Am J Hum Genet* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011 (2011).
- 20 46 Jia, X. *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One* **8**,  
21 e64683, doi:10.1371/journal.pone.0064683 (2013).
- 22 47 Pillai, N. E. *et al.* Predicting HLA alleles from high-resolution SNP data in three Southeast Asian  
23 populations. *Hum Mol Genet* **23**, 4443-4451, doi:10.1093/hmg/ddu149 (2014).
- 24 48 Genomes Project, C. *et al.* An integrated map of genetic variation from 1,092 human genomes.  
25 *Nature* **491**, 56-65, doi:10.1038/nature11632 (2012).
- 26 49 Tian, C. *et al.* Genome-wide association and HLA region fine-mapping studies identify  
27 susceptibility loci for multiple common infections. *Nat Commun* **8**, 599, doi:10.1038/s41467-017-  
28 00257-5 (2017).
- 29 50 Vilhjalmsón, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk  
30 Scores. *Am J Hum Genet* **97**, 576-592, doi:10.1016/j.ajhg.2015.09.001 (2015).
- 31 51 Khera, A. V. *et al.* Polygenic Prediction of Weight and Obesity Trajectories from Birth to  
32 Adulthood. *Cell* **177**, 587-596 e589, doi:10.1016/j.cell.2019.03.028 (2019).
- 33 52 Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with  
34 risk equivalent to monogenic mutations. *Nat Genet* **50**, 1219-1224, doi:10.1038/s41588-018-  
35 0183-z (2018).
- 36 53 Gorski, M. *et al.* 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function.  
37 *Sci Rep* **7**, 45040, doi:10.1038/srep45040 (2017).
- 38 54 Pattaro, C. *et al.* Genetic associations at 53 loci highlight cell types and biological pathways  
39 relevant for kidney function. *Nat Commun* **7**, 10023, doi:10.1038/ncomms10023 (2016).
- 40 55 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide  
41 association summary statistics. *Nat Genet* **47**, 1228-1235, doi:10.1038/ng.3404 (2015).
- 42 56 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature*  
43 **518**, 317-330, doi:10.1038/nature14248 (2015).
- 44 57 Heng, T. S., Painter, M. W. & Immunological Genome Project, C. The Immunological Genome  
45 Project: networks of gene expression in immune cells. *Nat Immunol* **9**, 1091-1094,  
46 doi:10.1038/ni1008-1091 (2008).
- 47 58 Eppig, J. T. *et al.* The Mouse Genome Database (MGD): facilitating mouse as a model for human  
48 biology and disease. *Nucleic Acids Res* **43**, D726-736, doi:10.1093/nar/gku967 (2015).

- 2 59 Bult, C. J. *et al.* Mouse genome database 2016. *Nucleic Acids Res* **44**, D840-847,  
3 doi:10.1093/nar/gkv1211 (2016).
- 4 60 Hormozdiari, F., Kostem, E., Kang, E. Y., Pasaniuc, B. & Eskin, E. Identifying causal variants at loci  
5 with multiple signals of association. *Genetics* **198**, 497-508, doi:10.1534/genetics.114.167908  
6 (2014).
- 7 61 Schmiedel, B. J. *et al.* Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression.  
8 *Cell* **175**, 1701-1715 e1716, doi:10.1016/j.cell.2018.10.022 (2018).
- 9 62 Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585,  
10 doi:10.1038/ng.2653 (2013).
- 11 63 Qiu, C. *et al.* Renal compartment-specific genetic variation analyses identify new pathways in  
12 chronic kidney disease. *Nat Med* **24**, 1721-1731, doi:10.1038/s41591-018-0194-4 (2018).
- 13 64 Suhre, K. *et al.* Human metabolic individuality in biomedical and pharmaceutical research. *Nature*  
14 **477**, 54-60, doi:10.1038/nature10354 (2011).
- 15 65 Shin, S. Y. *et al.* An atlas of genetic influences on human blood metabolites. *Nat Genet* **46**, 543-  
16 550, doi:10.1038/ng.2982 (2014).
- 17 66 Sun, B. B. *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73-79,  
18 doi:10.1038/s41586-018-0175-2 (2018).
- 19 67 Suhre, K. *et al.* Connecting genetic risk to disease end points through the human blood plasma  
20 proteome. *Nat Commun* **8**, 14357, doi:10.1038/ncomms14357 (2017).
- 21 68 Emilsson, V. *et al.* Co-regulatory networks of human serum proteins link genetics to disease.  
22 *Science* **361**, 769-773, doi:10.1126/science.aaq1327 (2018).
- 23 69 Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association  
24 studies using summary statistics. *PLoS Genet* **10**, e1004383, doi:10.1371/journal.pgen.1004383  
25 (2014).
- 26 70 Li, T. *et al.* A scored human protein-protein interaction network to catalyze genomic  
27 interpretation. *Nat Methods* **14**, 61-64, doi:10.1038/nmeth.4083 (2017).
- 28 71 Su, G., Kuchinsky, A., Morris, J. H., States, D. J. & Meng, F. GLayer: community structure analysis of  
29 biological networks. *Bioinformatics* **26**, 3135-3137, doi:10.1093/bioinformatics/btq596 (2010).
- 30 72 Nepusz, T., Yu, H. & Paccanaro, A. Detecting overlapping protein complexes in protein-protein  
31 interaction networks. *Nat Methods* **9**, 471-472, doi:10.1038/nmeth.1938 (2012).
- 32 73 Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular  
33 interaction networks. *Genome Res* **13**, 2498-2504, doi:10.1101/gr.1239303 (2003).
- 34 74 Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage,  
35 supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**,  
36 D607-D613, doi:10.1093/nar/gky1131 (2019).
- 37 75 Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment  
38 analysis and candidate gene prioritization. *Nucleic Acids Res* **37**, W305-311,  
39 doi:10.1093/nar/gkp427 (2009).
- 40 76 Graeber, T. G. & Eisenberg, D. Bioinformatic identification of potential autocrine signaling loops  
41 in cancers from gene expression profiles. *Nat Genet* **29**, 295-300, doi:10.1038/ng755 (2001).
- 42 77 Knox, C. *et al.* DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic*  
43 *Acids Res* **39**, D1035-1041, doi:10.1093/nar/gkq1126 (2011).
- 44 78 Zhu, F. *et al.* Therapeutic target database update 2012: a resource for facilitating target-oriented  
45 drug discovery. *Nucleic Acids Res* **40**, D1128-1136, doi:10.1093/nar/gkr797 (2012).
- 46 79 Carvalho-Silva, D. *et al.* Open Targets Platform: new developments and updates two years on.  
47 *Nucleic Acids Res* **47**, D1056-D1065, doi:10.1093/nar/gky1133 (2019).

- 2 80 Levey, A. S. *et al.* A new equation to estimate glomerular filtration rate. *Ann Intern Med* **150**, 604-  
3 612, doi:10.7326/0003-4819-150-9-200905050-00006 (2009).
- 4 81 Schwartz, G. J. *et al.* New equations to estimate GFR in children with CKD. *J Am Soc Nephrol* **20**,  
5 629-637, doi:10.1681/ASN.2008030287 (2009).
- 6 82 Stanaway, I. B. *et al.* The eMERGE genotype set of 83,717 subjects imputed to ~40 million  
7 variants genome wide and association with the herpes zoster medical record phenotype. *Genet*  
8 *Epidemiol* **43**, 63-81, doi:10.1002/gepi.22167 (2019).
- 9 83 Khan, A. *et al.* Medical Records-Based Genetic Studies of the Complement System. *J Am Soc*  
10 *Nephrol* **32**, 2031-2047, doi:10.1681/ASN.2020091371 (2021).
- 11 84 Shang, N. *et al.* Medical records-based chronic kidney disease phenotype for clinical care and "big  
12 data" observational and genetic studies. *NPJ Digit Med* **4**, 70, doi:10.1038/s41746-021-00428-1  
13 (2021).
- 14 85 McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*  
15 **48**, 1279-1283, doi:10.1038/ng.3643 (2016).
- 16 86 Abraham, G. & Inouye, M. Fast principal component analysis of large-scale genome-wide data.  
17 *PLoS One* **9**, e93766, doi:10.1371/journal.pone.0093766 (2014).
- 18 87 Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range  
19 of complex diseases of middle and old age. *PLoS Med* **12**, e1001779,  
20 doi:10.1371/journal.pmed.1001779 (2015).
- 21 88 Altshuler, D. M. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-+,  
22 doi:10.1038/nature15393 (2015).
- 23 89 Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate  
24 genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics*  
25 **44**, 955-+, doi:10.1038/ng.2354 (2012).
- 26 90 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**,  
27 203-+, doi:10.1038/s41586-018-0579-z (2018).
- 28 91 Denny, J. C. *et al.* PheWAS: demonstrating the feasibility of a phenome-wide scan to discover  
29 gene-disease associations. *Bioinformatics* **26**, 1205-1210, doi:10.1093/bioinformatics/btq126  
30 (2010).
- 31  
32