It is made available under a CC-BY 4.0 International license.

It is made available under a CC-BY 4.0 International license.

26 Abstract

27

28 Focal segmental glomerator (FOE) is a community temperator of any anti-
annual incidence in the United States in African-Americans compared to European-
ans of 24 cases and 5 cases per million, respectively. Among glomerul 29 and African-Americans. Therefore, a combined rare and common variant analysis was used to 30 Europe and Latin-America, FSGS was the second most frequent diagnosis, and in Asia the fifth.
We expand previous efforts in understanding genetics of FSGS by performing a case-control
study involving ethnically-diverse gro 31 Europe and previous efforts in understanding genetics of FSGS by performing a case-control
study involving ethnically-diverse groups FSGS cases (726) and a pool of controls (13,994),
using panel sequencing of approximately 32 Expand previous ethnically-diverse groups FSGS cases (726) and a pool of controls (13,994),
sting panel sequencing of approximately 2,500 podocyte-expressed genes. Through rare variant
association tests, we replicated kno 33 using panel sequencing of approximately 2,500 podocyte-expressed genes. Through rare variant
association tests, we replicated known risk genes - *KANK1, COL4A4,* and *APOL1*. A novel
significant association was observed f 34 association tests, we replicated known risk genes – *KANK1, COL4A4,* and *APOL1.* A novel
significant association was observed for the gene encoding complement receptor 1 (*CR1)*. High-
risk rare variants in *CR1* in the E 35 association tests, we replicated known risk genes – KANK1, CO24A4, and APOL1. A novel
significant association was observed for the gene encoding complement receptor 1 (CR1). High-
risk rare variants in CR1 in the European-36 significant association was observed for the gene encoding complement receptor 1 (*CR1*). High-

37 risk rare variants in *CR1* in the European-American cohort were commonly observed in Latin-

38 and African-Americans 37 risk rare variants in CR1 in the European-American cohort were commonity observed in Bathean
and African-Americans. Therefore, a combined rare and common variant analysis was used to
replicate the CR1 association in non-Eu 38 replicate the *CR1* association in non-European populations. The *CR1* risk variant, rs17047661, gives rise to the SI1/SI2 (R1601G) allele that was previously associated with protection against cerebral malaria. Pleiotrop 39 replicate the CR1 association in non-European populations. The CR1 risk variant, rs17047601,
gives rise to the S11/S12 (R1601G) allele that was previously associated with protection against
cerebral malaria. Pleiotropic ef 40 gives rise to the Sl₁/Sl₂ (R2) (R2) and that was previously associated with a gainst corebral malaria. Pleiotropic effects of rs17047661 may explain the difference in allele frequencies across continental ancestries an 41 cerebral materials are ontinental ancestries and suggest a possible role for genetically-driven
alterations of adaptive immunity in the pathogenesis of FSGS. 42 frequencies across continental and suppose and suppose and suppose and suppose and suppose $\frac{dS}{dt}$. 43 alterations of adaptive immunity in the pathogenesis of FSGS.

It is made available under a CC-BY 4.0 International license.

44 Introduction

45 For a mong both adults and children in the USA, and its incidence is increasing.^{1,2} The
ce and prevalence of FSGS are not precisely known due to the requirement of a kidney
for diagnosis and the lack of a central registr syndrome among both adults and children in the USA, and its incidence is increasing.^{1,2} The
incidence and prevalence of FSGS are not precisely known due to the requirement of a kidney
biopsy for diagnosis and the lack of 47 incidence and prevalence are and prevalence and prevalence and prevalence range from 1.4 to
21 cases per million population.³ The incidence of FSGS in the U.S. is about 4 times higher in
African Americans (6.8 patients p 48 50

between the U.S. is about 4 times higher in
African Americans (6.8 patients per million) and 2 times higher in Latin-Americans
(3.7 patients per million) compared to European-Americans (1.9 patients per million).⁴
Geneti 21 cases per million population.³ The incidence of FSGS in the U.S. is about 4 times higher in
50 African Americans (6.8 patients per million) and 2 times higher in Latin-Americans
51 (3.7 patients per million) compared (3.7 patients per million) compared to European-Americans (1.9 patients per million).⁴
Genetic studies of FSGS, conducted using both pedigree analyses and cohort-based
association studies, have identified a number of su (3.7 patients per million) compared to European-Americans (1.9 patients per million).⁴
Genetic studies of FSGS, conducted using both pedigree analyses and cohe
association studies, have identified a number of susceptibi 52 General Studies, have identified a number of susceptibility genes, yet explaining only a
of family-history enriched cases.⁵ The first genetic studies identifying the chromosome
2 region with FSGS were prompted by the ob 53 fraction of family-history enriched cases.⁵ The first genetic studies identifying the chromosome
(chr) 22 region with FSGS were prompted by the observed higher prevalence of FSGS in African
and African-American populati fraction of family-history enriched cases.⁵ The first genetic studies identifying the chromosome

for the 22 region with FSGS were prompted by the observed higher prevalence of FSGS in African

and African-American popu 55 The comparison of African-American populations, suggesting that one or more FSGS susceptibility gene

variants would be enriched on African-derived haplotypes.⁶⁻⁸ The subsequent discovery of

association of G1 and G2 cod 56 variants would be enriched on African-derived haplotypes.⁶⁻⁸ The subsequent discovery of association of G1 and G2 coding variants in *APOL1* with FSGS provided an explanation for increased prevalence of the disease amon variants would be enriched on African-derived haplotypes.^{6–8} The subsequent discovery of
58 association of G1 and G2 coding variants in *APOL1* with FSGS provided an explanation for
59 increased prevalence of the diseas 58 59 60

association of G1 and G2 coding variants in APOL1 with FSGS provided an explanation for
increased prevalence of the disease among African-descent populations. Pleiotropic properties
of these variants resulted in protection increase prevalence of the minimism and the diseased prevalence of the diseased FSGS risk.⁹
In this study, a large-scale genetic database was assembled from biopsy-confirmed
cases of Focal Segmental Glomerulosclerosis (F FSGS risk.⁹
In this study, a large-scale genetic database was assembled from biopsy-confirmed
cases of Focal Segmental Glomerulosclerosis (FSGS) and ethnically matched controls. The study
used a panel of approximately 61 FSGS risk.⁹
62 In
63 cases of Fo
64 used a par
65 role in the
66 study was
67 study is a :
68 and a more 62 Focal Segmental Glomerulosclerosis (FSGS) and ethnically matched controls. The study
panel of approximately \sim 2,500 genes associated with podocytes, which play a crucial
the formation and maintenance of the glomerular f 63 cases of a panel of approximately \sim 2,500 genes associated with podocytes, which play a crucial
role in the formation and maintenance of the glomerular filtration barrier. The purpose of the
study was to investigate the 64 used a panel of approximately $-2,500$ general and and policytes, which play a control in the formation and maintenance of the glomerular filtration barrier. The purpose of the study was to investigate the genetic basis o 65 role in the formation and maintenance of the glomerular filtration barrier. The purpose of study was to investigate the genetic basis of FSGS and identify novel susceptibility genes. This study is a significant extension o 66 study is a significant extension of previous work conducted by Yu et al⁵, with increased power
and a more diverse multi-ethnic cohort with greater sample size.
3 67 study is a significant extension of previous work conducted by Yu et al⁵, with increased power
68 and a more diverse multi-ethnic cohort with greater sample size.
69 68 and a more diverse multi-ethnic conductions are diverse multi-ethnic conductions and a more diverse multi-ethni
and a more diverse multi-ethnic conductions are diverse multi-ethnic conductions are diverse multi-ethnic cond

69

It is made available under a CC-BY 4.0 International license.

70

71 Methods

72 Confirmed FSGS⁶ and from patients diagnosed at Washington University. The research
Is were approved in advance and all subjects provided informed consent or assent. As
ples were de-identified, the Washington University i biopsy-confirmed FSGS⁶ and from patients diagnosed at Washington University. The research

protocols were approved in advance and all subjects provided informed consent or assent. As

all samples were de-identified, the 74 protocols were de-identified, the Washington University in St. Louis Institutional Review
Board (IRB) deemed that these studies did not require IRB approval. A total of 726 samples
were collected in a multicenter NIH study 75 Board (IRB) deemed that these studies did not require IRB approval. A total of 726 samples
were collected in a multicenter NIH study of biopsy-confirmed FSGS ⁶ and from patients
similarly diagnosed at Washington Universi 76 For collected in a multicenter NIH study of biopsy-confirmed FSGS ⁶ and from patients
similarly diagnosed at Washington University, the latter inherited from Yu et al ⁵. Genetic data
for cases were obtained using a "po were collected in a multicenter NIH study of biopsy-confirmed FSGS \circ and from patients

78 similarly diagnosed at Washington University, the latter inherited from Yu et al ⁵. Genetic data

79 for cases were obtained similarly diagnosed at Washington University, the latter inherited from Yu et al ⁵. Genetic data
for cases were obtained using a "podocyte exome" sequencing approach, consisting of a panel of
 2.482 genes, as described 79 2,482 genes, as described in Yu et al ⁵. Of the selected genes, mutations in five genes cause
familial FSGS and 200 genes are functionally related to these five. Additional genes were
selected based on expression profile 2,482 genes, as described in Yu et al ⁵. Of the selected genes, mutations in five genes cause

familial FSGS and 200 genes are functionally related to these five. Additional genes were

selected based on expression profi 81 faced based on expression profiles, as 677 genes are highly expressed in human micro-
dissected glomeruli, and 1600 genes are human orthologs of highly-expressed mouse podocyte
genes (**Figure 1A**). In the present study usi 82 83 85 86

discusser and J. In the present study using this panel, only the 2,482 genes constituting the
"podocyte exome" were analyzed, with 1.12% of the sequencing capture nucleotides located in
non-coding DNA.
The raw data files w genes (**Figure 1A**). In the present study using this panel, only the 2,482 genes constituting the
["]podocyte exome" were analyzed, with 1.12% of the sequencing capture nucleotides located in
⁸⁶non-coding DNA.
³⁷The ra The raw data files with sequencing reads (FASTQ files) were obtained for control
subjects from dbGAP general population cohorts not ascertained for kidney disease history
(**Sup. Table S1**). We extracted the regions sequenc The raw
subjects from db
(Sup. Table S1).
exome data of the
and 13,994 contr
construct a case
between cases an
calculated the fra 87 The rank and the requirements of the requirements of the radius of the rank of the radius of the rank files of the regions sequenced in the podocyte exome from the full-
data of the control cohort. There were 333,239 varia 88 (Sup. Table S1). We extracted the regions sequenced in the podocyte exome from the full-
exome data of the control cohort. There were 333,239 variants in the raw dataset of 726 cases
and 13,994 controls. We performed join 89 (Sup. Table S1). We extracted the regions sequenced in the podocyte exome from the full-
exome data of the control cohort. There were 333,239 variants in the raw dataset of 726 cases
and 13,994 controls. We performed jo 90 and 13,994 controls. We performed joint variant calling according to GATK best practices,¹⁰ to construct a case-control dataset. To confirm the absence of insufficient coverage biases between cases and controls, we cross and 13,994 controls. We performed joint variant calling according to GATK best practices,¹⁰ to

sonstruct a case-control dataset. To confirm the absence of insufficient coverage biases

between cases and controls, we cro 92 between cases and controls, we crossed the intervals common to both panels and then we
calculated the fraction of sequencing intervals that were well-covered (>10X) in cases and
controls; this was found to be 89% for both 93 calculated the fraction of sequencing intervals that were well-covered (>10X) in cases and controls, this was found to be 89% for both groups (**Figure S1**). In this calculation, only \overline{a} 94 controls; this was found to be 89% for both groups (**Figure S1**). In this calculation, only $\frac{4}{\sqrt{1-\frac{1}{2}}}$ 95 controls; this was found to be 89% for both groups (Figure S1). In this calculation, only

It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

variants that passed initial GATK hard-filtering¹⁰ were used. Next, the case-control dataset was
subjected to quality filtration using the Hail 0.2 open source software library ¹¹ (Figure S2).
The final data set includ subjected to quality filtration using the Hail 0.2 open source software library 11 (Figure S2).

The final data set included 577 cases (including 179 from Yu et al⁵), (including 378 fr

Yu et al⁵), and 131,179 var The final data set included 577 cases (including 179 from Yu et al⁵), (including 378 from

Yu et al⁵), and 131,179 variants. The drop-out rate was 60.64% for variants and 6.85% for

samples. The high drop-out rate for 99 Yu et al⁵), and 131,179 variants. The drop-out rate was 60.64% for variants and 6.85% for

90 samples. The high drop-out rate for variants is explained by exome sequencing being joined

91 with panel sequencing in a 100 101 102 103

samples. The high drop-out rate for variants is englanced by mini-organizing orang yindical
with panel sequences of controls and not sequenced in the case panel, due to broader DNA region
coverage in controls.
To account f exome sequences of controls and not sequenced in the case panel, due to broader DNA region
coverage in controls.
To account for population stratification, a joint principal components analysis (PCA)
was performed for case exore the controls.

To account for population stratification, a joint principal components analysis (PCA)

was performed for case and control genotypes. Uncorrelated common variants (linkage

disequilibrium pruning: r²< To account for
To account for
disequilibrium prunii
samples in PCA space
population stratificat
the genetic backgroun
First, we part
Clustering was perfo 104 To account for case and control genotypes. Uncorrelated common variants (linkage ibrium pruning: $r^2 < 0.9$; minor allele frequency - MAF>0.05) were used to cluster the in PCA space. To reduce the risk of false associatio 105 disequilibrium pruning: $r^2 < 0.9$; minor allele frequency – MAF>0.05) were used to cluster the
samples in PCA space. To reduce the risk of false associations in the rare variant analysis due to
population stratification, 107 108 109

disequilibrium pruning: $r^2<0.9$; minor allele frequency – MAF>0.05) were used to cluster the
samples in PCA space. To reduce the risk of false associations in the rare variant analysis due to
population stratification, samples in PCA space. To reduce the risk of and transfer false association stratification, we used principal components to subset the control cohort to match
the genetic background of cases.
First, we partitioned the datas First, we partitioned the dataset into clusters representing global population groups.
Clustering was performed using mixed Gaussian models AutoGMM package¹². The data were
stratified into eight clusters according to the First, we partitioned the
Clustering was performed using
stratified into eight clusters as
modeled by the AutoGMM algor
Project and labeled accordingly
South Asian and East Asian popu
due to small case count in each
retain 110 Figure 1.1 Matter into church of procession, protecting grown prepried to eight clusters according to the Gaussian mixture model. Agnostic clusters d by the AutoGMM algorithm were mapped to known clusters of the 1000 Genom Clustering was performed using mixed Gaussian models AutoGMM package¹². The data were
stratified into eight clusters according to the Gaussian mixture model. Agnostic clusters
modeled by the AutoGMM algorithm were mapped 112 stratified into eight) and algorithm were mapped to known clusters of the 1000 Genomes
Project and labeled accordingly (Figure S3). Two minor clusters of individuals belonging to
South Asian and East Asian populations and 113 modeled accordingly **(Figure S3)**. Two minor clusters of individuals belonging to South Asian and East Asian populations and admixed ancestry were excluded from the analysis due to small case count in each cluster, resulti 114 Project and labeled accordingly (Figure 33). Two minor clusters of individuals belonging to South Asian and East Asian populations and admixed ancestry were excluded from the analysis due to small case count in each cluste 115 South Asian and East Asian preparations and anti-dimensional term in an admised to small case count in each cluster, resulting in low statistical power. Of the six clusters retained for further study, three included indivi 116 retained for further study, three included individuals of European descent and reflected
different local-population origins; these minor-clusters were combined into a single major-
European cluster. The fourth cluster repr 117 retained into a single major-
European cluster. The fourth cluster represented the individuals with African ancestry. Two
other clusters belonged to the Latin-American population. After filtering out the low power
clusters 118 119 other clusters belonged to the Latin-American population. After filtering out the low power
clusters, the dataset had 551 cases and 11,591 controls (**Figure S4**).
5 120 clusters, the dataset had 551 cases and 11,591 controls (**Figure S4**).
5 121 clusters, the dataset had 551 cases and 11,591 controls (Figure S4).

It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

Further case-control matching was conducted using the MatchIt¹³ package (**Figure 1B,**
 Figure S4). 16 cases were excluded from further consideration because it was not possible to

select appropriate population contro 123 Figure S4). 16 cases were excluded from further consideration because it was not possible to select appropriate population controls for them. The final dataset included 358 cases and 1,488 controls for the European-Ame 124 select approach the European-American cluster, 125 cases and 137 controls for the African-American cluster and 52 cases and 288 controls for the Latin-American cluster (Figure S4). The Weir and Cockerham F-statistic for an 125 significant power increase could not be achieved with a larger number of controls (Figure $S5$). 126 Merican cluster and 52 cases and 288 controls for the Latin-American cluster (Figure S4). The
Weir and Gockerham F-statistic for analysis of population structure showed that the European-
American cluster and the African-A 127 those of similar studies on European cohorts, but they may not be sufficient to detect genome-128 (weighted mean fixation index Fst=0.0878 for case cohorts), demonstrating distinct population
differences. The Latin-American cluster is also sufficiently isolated from the other clusters
(weighted mean fixation index Fst= 129 of the Unitaryton index Fst=0.0148). The power analysis for the European-American (weighted mean fixation index Fst=0.0148). The power analysis for the European-American dataset showed many-fold power superiority over prev 130 (weighted mean fixation index Fst=0.0148). The power analysis for the European-American dataset showed many-fold power superiority over previous FSGS cohort studies ¹⁴, and that an exponential power increase threshold ha 131 (weighted mean finally first since μ). The power index FSGS cohort studies ¹⁴, and that an exponential power increase threshold has been reached with the current number of cases, *i.e.*, a significant power increase dataset showed many-fold power superiority over previous FSGS cohort studies ¹⁴, and that an exponential power increase threshold has been reached with the current number of cases, *i.e.*, a significant power increase c 133 exponential power increase threshold has been reached with the current number of cases, i.e., a
significant power increase could not be achieved with a larger number of controls (Figure S5).
The values of significant powe 135 136 137

134 significant power increase could not be achieved with a larger number of controls (Figure S5).

135 The values of significant power for the African and Latin-American clusters are comparable to

136 those of similar s The values of similar studies on European cohorts, but they may not be sufficient to detect genome-
wide significant associations of frequent variants because of their small effect size (**Figure S6**).
We performed an asso We performed an association study using common synonymous variants (gnomAI
population specific $AF \ge 0.01$), as these are unlikely to contribute to a phenotype and yet reflec
possible ancestral bias between case and matche 138 Francestral bias between case and matched controls cohorts. For all three European-
an, African-American and Latin-American post-matching datasets we confirmed the
of systematic bias between cases and controls (**Figure 1C** 139 possible ancestral bias between case and matched controls cohorts. For all three European-
American, African-American and Latin-American post-matching datasets we confirmed the
absence of systematic bias between cases and 140 141 142

Using a data set of matched cases and controls after quality filtration, we performed American-American and Controls (Figure 1C).

Acsults

Using a data set of matched cases and controls after quality filtration, we performed

several association studies. Because analyses of this dataset were limited to the absence of systematic bias between cases and controls (Figure 1C).

Results

Using a data set of matched cases and controls after quali

several association studies. Because analyses of this dataset wer

exome", we focused 143 Results
144 l
145 several
146 exome", 144 association studies. Because analyses of this dataset were limited to the "podocyte
we focused on missense variants and protein truncation variants (PTV).
6 145 sexome", we focused on missense variants and protein truncation variants (PTV).

6 146 exome", we focused on missense variants and protein truncation variants (PTV).

It is made available under a CC-BY 4.0 International license.

147 For each conductional covariates. 3,777 variants with missense and PTV (stop_gained, ift_variant, splice_acceptor_variant, splice_donor_variant) effects on protein were in this analysis. In the European-American cluster, 148 frameshift_variant, splice_acceptor_variant, splice_donor_variant) effects on protein were
included in this analysis. In the European-American cluster, two variants were significantly
associated with FSGS: (1) rs601314 – 149 included in this analysis. In the European-American cluster, two variants were significantly
associated with FSGS: (1) $rs601314 - p=8.1x10^{-9}$, reference allele is a minor allele;
 $OR_{minor} allele = 13.24$ (Cl_{95%}=[3.996, 56.51] 150 associated with FSGS: (1) rs601314 - p=8.1x10⁻⁹, reference allele is a minor allele;

OR_{minorallele}=13.24 (Cl_{95%}=[3.996, 56.51]), missense, *EFEMP2* (EGF-containing fibulin

extracellular matrix protein 2); and (2) 151 associated with FSGS: (1) rs601314 – p=8.1x10⁻⁹, reference allele is a minor allele;

152 OR_{minorallet}=13.24 (Cl_{95%}=[3.996, 56.51]), missense, *EFEMP2* (EGF-containing fibulin

153 extracellular matrix protein 2 152 OR_{minor allele}=13.24 (CI_{95%}=[3.996, 56.51]), missense, *EFEMP2* (EGF-containing fibulin extracellular matrix protein 2); and (2) rs117071588 – p=4.0x10⁻⁶, alternative allele is a minor allele; OR_{minor allele}; extracellular matrix protein 2); and (2) rs117071588 – p=4.0x10⁻⁶, alternative allele is a minor

allele; OR_{minoraliet}=11.66 (Cl_{95%}=[2.790, 68.35]), missense, *CCDC82* (coiled-coiled domain

containing 82), Signific 154a allele; OR_{minor allele}=11.66 (CI_{95%}=[2.790, 68.35]), missense, *CCDC82* (coiled-coiled domain containing 82), Significance threshold was determined with Bonferroni correction - $p<0.05/3,777=1.32x10^{-5}$; **Figure** 155 containing 82), Significance threshold was determined with Existential contains people.

in the African-American cohort showed that the rs601314 variant (*EFFMP2*) was not

significantly associated with FSGS (p=0.062, ref 156 p<0.05/3,777=1.32x10⁻⁵; **Figure S7, Sup. Table S2**). Replication analysis of these two variants

157 in the African-American cohort showed that the rs601314 variant (*EFFMP2*) was not

158 significantly associated w 157 in the African-American cohort showed that the rsoot of a minor allele,

oR_{minor allele}, -0.7418 , $C1_{95\%}$ =[0.5370, 1.015]) and that rs117071588 (*CCDC82*) was absent from

the African-American dataset. A similar si 158 Significantly associated with FSGS (pseudomatric minimum alleless). The minimum alleless (RCDC82) was absent from
the African-American dataset. A similar situation was observed when replication was attempted
in the Latin-159 OR_{minor allele}=0.7418, Cl_{95%}=[0.5370, 1.015]) and that rs117071588 (*CCDC82*) was absent from

160 the African-American dataset. A similar situation was observed when replication was attempted

161 in the Latin-Am 160 the Latin-American cohort: rs601314 was not significantly associated (p=0.097, reference
allele is a minor allele, OR_{minor allele}=2.99, Cl_{95%}=[0.6891, 14.76]), rs117071588 was absent from
the data. Despite the signific 161 in the Latin-American contract terms in the Latin-Agaminality attention (p=0.000) and defined at a. Despite the significant association statistics of *EFEMP2* and *CCDC82* in the European-American group, the lack of replic 162 allele is a minor allele is a minor allele is a minor allele is a minor allele is a meta-
American group, the lack of replication of the *EFEMP2* variant FSGS association in the other
populations makes this a less robust f 163 164 165 166

American group, the lack of replication statistics of *EFEMP2* variant FSGS association in the other
populations makes this a less robust finding but might serve as a starting point for future
studies.
The rs601314 in *EFE* Experient group, the lack of replication of the EFEMP2 variant FSGS association in the other
populations makes this a less robust finding but might serve as a starting point for future
studies.
The rs601314 in *EFEMP2* (NC populations.

The rs601314 in *EFEMP2* (NC_000011.9:g.65636053T>C, ENSP00000434151:p.1259V)

and rs117071588 in *CCDC82* (NC_000011.9:g.96117537A>C, ENSP00000278520:p.D125E)

variants are defined by most common *in silico* standers
and rs1
variants
classifie
of misse
Variant
variant 167 The rs601314 in EFEMP2 (NC_000011.9:g.056366351>C, ENSP00000278520:p.D125E)

117071588 in *CCDC82* (NC_000011.9:g.96117537A>C, ENSP00000278520:p.D125E)

s are defined by most common *in silico* pathogenicity predictors as 168 and 13117071580 in CCDC02 (NC_000011.7:g.96117537A>C, ENSP00000278520:p.D125E)
variants are defined by most common *in silico* pathogenicity predictors as benign ¹⁵. FATHMM
classified rs601314 as "damaging" (Fathmm Scor variants are defined by most common *in silico* pathogenicity predictors as benign ¹⁵. FATHMM

classified rs601314 as "damaging" (Fathmm Score Converted = 0.48) ¹⁵. The specific predictors

of missense deleteriousness classified rs601314 as "damaging" (Fathmm Score Converted = 0.48) ¹⁵. The specific predictors
171 of missense deleteriousness classified rs601314 and rs117071588 as benign (MISTIC<0.5) ¹⁶.
172 Variant rs601314 affects of missense deleteriousness classified rs601314 and rs117071588 as benign (MISTIC<0.5) ¹⁶.
172 Variant rs601314 affects the von Willebrand factor type A (vWA) domain of *EFFMP2* and
173 variant rs117071588 affects the d 172 variant rs117071588 affects the domain of unknown function (DUF4196) of $CCDC82$ ¹⁵. Both
7 173 variant rs117071588 affects the domain of unknown function (DUF4196) of *CCDC82* ¹⁵. Both
7

It is made available under a CC-BY 4.0 International license.

174

variants have a missense control on present burden analyses in the European-American cohort, focusing
on missense variants and PTV with a population frequency below 0.01. This cutoff was chosen
according to the presence of proteins (MPC<2.0) ¹⁷.

176 We carried out

177 on missense variants a

178 according to the prese

179 the interval [0; 0.01] (**I**

180 five tests representing

181 all potential risk patte

182 burden tests are useful
 176 The variance variant burden and analyses in the European-American carry colling

ense variants and PTV with a population frequency below 0.01. This cutoff was chosen

ng to the presence of a signal in each of the quartiles 177 according to the presence of a signal in each of the quartiles of allele frequency distribution in
the interval [0; 0.01] (Figure S8). A rare variant association study (RVAS) was performed using
five tests representing dif 178 according to the present of the presence of methods for each gene, in order to identify
five tests representing different statistical classes of methods for each gene, in order to identify
all potential risk patterns. If m 179 the interval [0; 0.01] (**Figure S8**). A rare variant association study (RVAS) was performed using
180 five tests representing different statistical classes of methods for each gene, in order to identify
181 all potenti 180 all potential risk patterns. If most variants are causal and have unidirectional effects, classical
burden tests are useful, due to their high power. However, adaptive burden tests are considered
more reliable than those u 181 all potential rate patterns. If an external and the most variant restant rests are considered
more reliable than those using fixed weights or thresholds. ¹⁸ In addition, some tests can
improve understanding of the result 182 burden tests are using fixed weights or thresholds.¹⁸ In addition, some tests can
improve understanding of the results. Tests of variance components are effective when there
are variants that both increase and decrease a more reliable than those using fixed weights or thresholds. ¹⁸ In addition, some tests can

184 improve understanding of the results. Tests of variance components are effective when there

185 are variants that both incr 184 ¹⁸. These tests included Fisher's exact test, C-alpha, adaptive sum statistic (ASUM), weighted
¹⁸. These tests included Fisher's exact test, C-alpha, adaptive sum statistic (ASUM), weighted
sum statistics (WSS) and ker 185 ¹⁸. These tests included Fisher's exact test, C-alpha, adaptive sum statistic (ASUM), weighted
sum statistics (WSS) and kernel-based adaptive clustering (KBAC) (**Figure S9**). The resulting p-
values were combined using 186 ¹⁸. These tests included Fisher's exact test, C-alpha, adaptive sum statistic (ASUM), weighted im statistics (WSS) and kernel-based adaptive clustering (KBAC) (**Figure S9**). The resulting p-
alues were combined using the Simes method for multiple hypothesis testing, which is suitable
r merging dependent test statis 187 sum statistics (WSS) and kerner-based adaptive clustering (KDAC) (Figure 35). The resulting p-
values were combined using the Simes method for multiple hypothesis testing, which is suitable
for merging dependent test stati 188 For merging dependent test statistics (**Sup. Table S3**). The top 10 associated genes included
four genes, *APOL1*, *KANK1*, *COL4A4*, *IL36G*, that were previously identified in FSGS association
studies. Of these, the top 189 four genes, *APOL1*, *KANK1*, *COLAAA*, *IL36G*, that were previously identified in FSGS association
studies. Of these, the top two genes reached significance after Bonferroni correction
(p=0.05/2,482=2.015x10⁻⁵): *APOL1* 190 191

four genes, APOL1, KANK1, COLFA4, ILS00, that were previously identified in F5G5 association
studies. Of these, the top two genes reached significance after Bonferroni correction
(p=0.05/2,482=2.015x10⁻⁵): *APOL1* (p=1.4 (p=0.05/2,482=2.015x10⁻⁵): *APOL1* (p=1.47x10⁻⁶), a known FSGS susceptibility gene, and *CR1*
(p=1.67x10⁻⁵), a novel candidate gene (**Figure 2A, Sup. Figure S10**).
Significantly-associated genes in the European-Amer (p=0.05/2,482=2.015x10⁻⁵): *APOL1* (p=1.47x10⁻⁶), a known FSGS susceptibility gene, and *CR1*

(p=1.67x10⁻⁵), a novel candidate gene (**Figure 2A, Sup. Figure S10**).

Significantly-associated genes in the European-Am 193 (p=1.67x10-5), a novel candidate gene (**Figure 2A, Sup. Figure S10**).

194 Significantly-associated genes in the European-American coh

195 in a replication study of the African-American and Latin-American c

196 *CR1* 194 Significantly intertained generation and Latin-American cohorts. Neither *APOL1* nor
increase replicated using rare (MAF<0.01) variant analysis (**Sup. Table S4**). Previously-
d positive selection acting on the *APOL1*⁹ v 195 In a replication study of the African-American and Latin-American cohorts. Neither Afronz thor
CR1 were replicated using rare (MAF<0.01) variants in the African-American population
suggests that FSGS risk variants might be 196 CR1 were replicated using rare (MAF<0.01) variant analysis (**Sup. Table S4**). Previously-

197 observed positive selection acting on the $APOL1^9$ variants in the African-American population

198 suggests that FSGS risk observed positive selection acting on the *APOL19* variants in the African-American population

198 suggests that FSGS risk variants might be too common to be detected by a RVAS in non-

European populations. Therefore, w 198 199 200 association signals in $APOL1$ and CR1.

It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

201 In APOL1 and CR1 and four variants, consisting of pair-locus G1 in APOL1 (rs60910145

2022.10:g.36662034T>G, ENSP00000317674.4:p.lle400Met] and rs73885319

2022.10:g.36661906A>G, ENSP00000317674.4:p.Ser358Gly]⁹) and two 202 signals in APOL1 and CR1 and Four variants, consisting of pair-locus G1 in APOL1 (1500 210145)

[NC_000022.10:g.36661906A>G, ENSP00000317674.4:p.Ser358Gly]⁹) and two closely-adjacent

variants in CR1 (rs17047661 [NC_000 203 [NC_000022.10:g.36661906A>G, ENSP00000317674.4:p.Ser358Gly]⁹) and two closely-adjacent
variants in $CR1$ [rs17047661 [NC_000001.10:g.207782889A>G,
ENSP00000356016.4:p.Lys2040Glu]) were selected for replication in other a 204 [NC_000022.10:g.36661906A>G, ENSP00000317674.4:p.Ser358Gly]⁹) and two closely-adjacent

205 variants in CRI (rs17047661 [NC_000001.10:g.207782889A>G,

206 ENSP00000356016.4:p.Arg2051Gly] and rs17047660 [NC_000001.10 205 ENSP00000356016.4:p.Arg2051Gly] and rs17047660 [NC_000001.10:g.207782856A>G,

ENSP00000356016.4:p.Arg2051Gly] were selected for replication in other ancestries

(Figure 2B). Additionally, we eliminated the possibility of 206 ENSP00000356016.4:p.Lys2040Glu]) were selected for replication in other ancestries

(Figure 2B). Additionally, we eliminated the possibility of this result being a false positive due

to coverage imbalance in the associat 207 ENSP00000356016.4:p.Lys2040Glu]) were selected for replication in other ancestries **208 (Figure 2B)**. Additionally, we eliminated the possibility of this result being a false positive due

209 to coverage imbalance in the associated genes (**Sup. Figure S11**). *APOL1* variants were

210 successfully repl 209 successfully replicated (rs73885319: $p=0.001779$, alternative allele is a minor allele,
OR_{minorallele}=1.59, Cl_{95%}=[1.18, 2.14]; rs60910145: $p=0.002271$, alternative allele is a minor allele,
oR_{minorallele}=1.59, Cl 210 successfully replied in the state of the state of the original European-American
allele, OR_{minorallele}=1.58, Cl_{95%}=[1.17, 2.12]). The variants in *CR1* were more common and had
smaller effect size, therefore, we lacked 211 OR_{minor allele}=1.59, Cl_{95%}=[1.18, 2.14]; rs60910145: p=0.002271, alternative allele is a minor allele, OR_{minor allele} =1.58, Cl_{95%}=[1.17, 2.12]). The variants in *CR1* were more common and had smaller effect s 213 214 215

212a allele, OR_{minor allele} =1.58, Cl_{95%}=[1.17, 2.12]). The variants in *CR1* were more common and had

213a smaller effect size, therefore, we lacked the statistical power to see the significant replication

214a (rs frs17047661: p=0.28, reference allele is a minor allele, $OR_{minor\,allet} = 1.22$, $C_{155\%} = [0.84, 1.76]$;
rs17047660: p=0.74, alternative allele is a minor allele, $OR_{minor\,allet} = 1.09$, $C_{155\%} = [0.69, 1.71]$).
Second replicatio (rs17047660: p=0.74, alternative allele is a minor allele, $OR_{\text{minor allele}} = 1.09$, $C_{\text{loss}} = [0.69, 1.71]$).
Second replication was attempted in the Latin-American cohort because the variant
frequencies for the variants of int rstate also, externative allele is a minister alleled in the Latin-American cohort because the vari-
frequencies for the variants of interest are more similar to the original European-Americ
cohort. Variants rs17047661 in 216 Second replication was attempted in the Latin-American cohordinal European-American
Variants rs17047661 in CR1 and rs60910145 and rs73885319 in *APOL1* surpassed the
ion significance threshold (p=0.05/4=0.0125) (**Sup. Tab** 217 frequencies for the variants for the variants for the variants of the replication significance threshold (p=0.05/4=0.0125) (Sup. Table S5). Analysis of the statistical power for identified effect sizes in Latin-American an 218 cohort. Variants rs17047601 in ch1 and r500910145 and rs73065512 in APOL1 surpassed the
replication significance threshold (p=0.05/4=0.0125) (**Sup. Table S5**). Analysis of the
statistical power for identified effect sizes 219 220 221 222

replication significance threshold (p=0.057+=0.0125) (Sup. Table S5). Analysis of the
statistical power for identified effect sizes in Latin-American and African-American cohorts
indicated that the lack of replication in indicated that the lack of replication in the latter is most likely driven by the statistical power

limitations (Sup. Figure S6).

Meta-estimates of pMETAL¹⁸ for all 4 variants were also calculated for the European-Ame limitations (Sup. Figure S6).

Meta-estimates of pMETAL¹⁸ for all 4 variants were also calculated for the European-American and Latin-American cohorts: rs60910145 ($p_{META} = 9.706 \times 10^{-6}$), rs73885319

($p_{META} = 1.420 \times 1$ Meta-estimates of pM
Meta-estimates of pM
American and Latin-Amer
(PMETAL=1.420x10⁻⁴), rs170
Interestingly, the variants in
of four possible haplotyp 223 Meta-estimates of pMETAL ¹⁸ for all 4 variants were also calculated for the European-

224 American and Latin-American cohorts: rs60910145 ($p_{META} = 9.706 \times 10^{-6}$), rs73885319

225 ($p_{META} = 1.420 \times 10^{-4}$), rs170476 224 American and Latin-American cohorts: rs60910145 ($p_{META} = 9.706 \times 10^{-6}$), rs73885319
225 ($p_{META} = 1.420 \times 10^{-4}$), rs17047660 ($p_{META} = 0.6359$), rs17047661 ($p_{META} = 9.314 \times 10^{-3}$).
226 Interestingly, the variants in 225 ($p_{METAL}=1.420x10^{-4}$), rs17047660 ($p_{METAL}=0.6359$), rs17047661 ($p_{METAL}=9.314x10^{-3}$).
226 Interestingly, the variants in *CR1*: rs17047661 and rs17047660 are linked with only three out
227 of four possible haplotypes 226 Interestingly, the variants in CR1: rs17047661 and rs17047660 are linked with only three out
of four possible haplotypes observed in African subpopulations in 1000 genomes
9 227

9 $\begin{array}{ccc} \hline \end{array}$ of four possible happens in $\begin{array}{ccc} \hline \end{array}$

It is made available under a CC-BY 4.0 International license.

228 (AFR:YRI+LWK+GWD+MSL+ESN+ASW+ACB: $r^2=0.15$, $D'=1$; YRI: $r^2=0.15$, $D'=1$;
229 ASW: $r^2=0.18$, $D'=1$; ACB: $r^2=0.13$, $D'=1$) and observed in global Latin-American population
230 (AMR:MXL+PUR+CLM+PEL: $r^2=0.46$, 230 ASW: $r^2=0.18$, D'=1; ACB: $r^2=0.13$, D'=1) and observed in global Latin-American population

230 (AMR:MXL+PUR+CLM+PEL: $r^2=0.46$, D'=1).

231 Next, the normalized integral haplotype score (iHS) was directly estima 230 $(AMR:MXL+PUR+CLM+PEL: r^2=0.46, D'=1).$

231 Next, the normalized integral hap

232 discovery cohort for the variants included i

233 cohort, selection pressure analysis confirm

234 alleles: rs73885319 (iHS=-2.16), rs609101
 231 Next, the normalized imagina happed point of the normal in the more proportion pressure analysis confirmed positive selection (iHS<-2.0) for the G1 APOL1 rs73885319 (iHS=-2.16), rs60910145 (iHS=-2.21) and revealed positive 232 discovery content in the variance matrix in the vary content analysis. The G1 AP0L1
alleles: rs73885319 (iHS=-2.16), rs60910145 (iHS=-2.21) and revealed positive selection for
rs17047660 (iHS=-2.71) in CR1, whereas no such 233 234 235 236 237

alleles: rs73885319 (iHS=-2.16), rs60910145 (iHS=-2.21) and revealed positive selection for
rs17047660 (iHS=-2.71) in CR1, whereas no such selection was detected for rs17047661 (iHS=-
1.04). The following results were obt alleles: rs17047660 (iHS=-2.71) in CR1, whereas no such selection was detected for rs17047661 (iHS=-1.04). The following results were obtained for the Latin American cohort: rs73885319 (iHS=-1.99), rs60910145 (iHS=-2.01), 1.04). The following results were obtained for the Latin American cohort: rs73885319 (iHS=-1.99), rs60910145 (iHS=-2.01), rs17047660 (iHS=-0.142) and rs17047661 iHS=1.13).

Allele frequencies for all variants included in 1.99), rs60910145 (iHS=-2.01), rs17047660 (iHS=-0.142) and rs17047661 iHS=1.13).

Allele frequencies for all variants included in the replication analyses are significantly

different between population groups, which can Allele frequencies for all variants included in the replication analyses are sig
different between population groups, which can nominally indicate either positive se
genetic drift. These included the following variants: r 238 Allele frequencies for all variables for all variables for all variables included the following variants: rs60910145: gnomAD EUR AF=8.6x10.5,

DAFR AF=0.23, $p=2.2x10^{16}$; rs73885319: gnomAD EUR AF=1.1x10⁴,

DAFR AF=0. 239 matrix of the included the following variants: $rs60910145$: gnomAD EUR AF=8.6x10⁻⁵,
gnomAD AFR AF=0.23, $p=2.2x10^{-16}$; $rs73885319$: gnomAD EUR AF=1.1x10⁻⁴,
gnomAD AFR AF=0.23, $p=2.2x10^{-16}$; $rs17047661$: gnomAD EUR genetic drift. These included the following variants: rs60910145: gnomAD EUR AF=8.6x10⁻⁵,

gnomAD AFR AF=0.23, p=2.2x10⁻¹⁶; rs73885319: gnomAD EUR AF=1.1x10⁻⁴,

gnomAD AFR AF=0.23, p=2.2x10⁻¹⁶; and rs17047661: gno 241 gnomAD AFR AF=0.23, p=2.2x10⁻¹⁶; rs73885319: gnomAD EUR AF=1.1x10⁻⁴,
242 gnomAD AFR AF=0.23, p=2.2x10⁻¹⁶; rs17047661: gnomAD EUR AF=3.0x10⁻³,
243 gnomAD AFR AF=0.24, p=2.2x10⁻¹⁶; and rs17047660: gnomAD EUR A 242 gnomAD AFR AF=0.23, p=2.2x10⁻¹⁶; rs17047661: gnomAD EUR AF=3.0x10⁻³,
243 gnomAD AFR AF=0.62, p=2.2x10⁻¹⁶; and rs17047660: gnomAD EUR AF=1.0x10⁻³,
244 gnomAD AFR AF=0.24, p=2.2x10⁻¹⁶). It is likely that iHS e 243 gnomAD AFR AF=0.62, p=2.2x10⁻¹⁶; and rs17047660: gnomAD EUR AF=1.0x10⁻³,
244 gnomAD AFR AF=0.24, p=2.2x10⁻¹⁶). It is likely that iHS estimates can be skewed by complex
245 population structure or demographic var gnomAD AFR AF=0.24, p=2.2x10⁻¹⁶). It is likely that iHS estimates can be skewed by complex
population structure or demographic variables such as population growth, bottleneck events,
and changes in recombination and mut 245 population structure or demonstration structure or demonstration structure or demonstration and African-American populations in 1000 genomes significantly vary for rs17047661 (AF_{YRI}=0.69, AF_{iWK}=0.70, AF_{6WD}=0.79, AF₅ 246 247 248

and mutation and African-American populations in 1000 genomes significantly vary for rs17047661 (AF_{YN}=0.69, AF_{LWK}=0.70, AF_{WW}=0.79, AF_{NSL}=0.79, AF_{ESN}=0.72, AF_{ASW}=0.58, AF_{ACB}=0.66).
AF_{ACB}=0.66).
We sought evi African-African-American and African-African-African-African-African-African-African-African-African-African-American case cohort and compared this with the expectation of random assortment. There are cohordered in the Af RF_{ACB}=0.66).
We sought evidence of co-evolving changes in allele frequencies between (a) the G1 and
G2 variants in *APOL1* and (b) replicated rs17047661 in *CR1*. We estimated the number of
individuals who carry both rs1 249 AF_{ACB}=0.66).
250 We so
251 G2 variants
252 individuals w
253 American ca. 250 ants in $APOL1$ and (b) replicated rs17047661 in CR1. We estimated the number of als who carry both rs17047661 and either one or both G1 and G2 alleles in the African-
an case cohort and compared this with the expectation o 251 G2 variants in APOL1 and (b) replicated rs17047661 in CR1. We estimated the number of individuals who carry both rs17047661 and either one or both G1 and G2 alleles in the African-American case cohort and compared this wi 252 253 $\frac{1}{2}$ and compared the expectation of random association of ran

It is made available under a CC-BY 4.0 International license.

254 255

effects of rs17047661 are fully independent of those of *APOL1*.

Most common in silico variant effect predictors do not categorize rs17047661

(NC_000001.10:g.207782889A>G, ENSP00000383744:p.R1601G) as pathogenic ¹⁵. Ex Most common in silico variant effect predictors (NC_000001.10:g.207782889A>G, ENSP00000383744:p.R160:
are, for example, PolyPhen2 HVAR and MutationAssessor,
"probably damaging" (Polyphen 2 Hvar Score = 0.964) and
(Mutation 256 Most Common in Same Categoria in the Predictor of the Basily Control of the Secreptions

Example, PolyPhen2 HVAR and MutationAssessor, which classify rs17047661 as

by damaging" (Polyphen 2 Hvar Score = 0.964) and "medium 267 (NC_000001.10:g.207782889A>G, ENSP00000383744:p.R1601G) as pathogenic ¹⁵. Exceptions
258 are, for example, PolyPhen2 HVAR and MutationAssessor, which classify rs17047661 as
²⁵⁹ "probably damaging" (Polyphen 2 Hvar 258 are, for example, Polyphen2 Hvare Score = 0.964) and "medium functional effect"

(Mutationassessor Score Converted = 0.70), respectively ¹⁵. The rs17047661 variant of CR1

induces a missense effect (p.R1601G) in the pro 259 (Mutationassessor Score Converted = 0.70), respectively ¹⁵. The rs17047661 variant of CR1 induces a missense effect (p.R1601G) in the protein domains common to secreted complement fixation protein (PHA02927) and complem 260 (Mutationassessor Score Converted = 0.70), respectively ¹⁵. The rs17047661 variant of CR1 induces a missense effect (p.R1601G) in the protein domains common to secreted complement fixation protein (PHA02927) and com 261 fixation protein (PHA02927) and complement control protein (CCP) modules, also known as
consensus short repeats (SCR) or SUSHI repeats. Specifically, it affects the one of the four Long
Homologous Repeats (LHRs), LHR-D, w 262 consensus short repeats (SCR) or SUSHI repeats. Specifically, it affects the one of the four Long

Homologous Repeats (LHRs), LHR-D, which is responsible for binding C1q, Mannose Binding

lectin (MBL) and ficolin ^{15,19}. 263 Homologous Repeats (LHRs), LHR-D, which is responsible for binding C1q, Mannose Binding
lectin (MBL) and ficolin ^{15,19}. However, the specific predictors of missense deleteriousness
classified rs17047661 as benign (MISTIC 264 Homologous Repeation (MBL) and ficolin 15.19. However, the specific predictors of missense deleteriousness
classified rs17047661 as benign (MISTIC<0.5) ¹⁶, which is likely due to the nonconservative
nature of the affecte

268

269 Discussion

lectin (MBL) and ficolin ^{15,19}. However, the specific predictors of missense deleteriousness

266. classified rs17047661 as benign (MISTIC<0.5) ¹⁶, which is likely due to the nonconservative

267. nature of the affecte classified rs17047661 as benign (MISTIC<0.5) ¹⁶, which is likely due to the nonconservative

267 nature of the affected region (MPC<2.0) ¹⁷.

268

269 Discussion

270 The complement system is a complex network of prote 267 nature of the affected region (MPC<2.0) ¹⁷.

268 Discussion

270 The complement system is a compl

271 protecting the body against microbial is

272 classical immune pathway in response to l

273 alternative pathway 270 The complete system in the body against microbial infections, which are activated either through the

l immune pathway in response to binding to Fc-fragments of lgM or lgG, or through an

ive pathway of non-specific bindin 271 protecting the body against microbial interacting the body and classical immune pathway of non-specific binding to antigens on membranes or to mannose residues through the lectin pathway.²⁰ Each protein in the cascade is 272 alternative pathway of non-specific binding to antigens on membranes or to mannose residues
through the lectin pathway.²⁰ Each protein in the cascade is activated by proteolysis, splitting
the original proenzyme into "a" 273 alternative parameters of non-specific binding the original proenzyme into "a" and "b" structures (the exception is C1, which splits into q, r, s molecules). The large molecule "b" is directly involved in the sequential ac 274 through the lectin pathway.²⁰ Each protein in the cascade is activated by proteolysis, splitting
275 the original proenzyme into "a" and "b" structures (the exception is C1, which splits into q, r, s
276 molecules). 275 the original probability in the sequential activation of the complement system and the small molecule "a" is an anaphylatoxin, which causes degranulation of mast cells and chemotaxis of other immune cells such as neutrophi 276 complement system and the small molecule "a" is an anaphylatoxin, which causes degranulation
of mast cells and chemotaxis of other immune cells such as neutrophils, eosinophils, monocytes,
and T lymphocytes. All of these f 277 of mast cells and chemotaxis of other immune cells such as neutrophils, eosinophils, monocytes,
and T lymphocytes. All of these factors have the potential to contribute to either innate immune
11 278 and T lymphocytes. All of these factors have the potential to contribute to either innate immune 11 279 and T lymphocytes. All of these factors have the potential to either inner inner inner innate immune to either inner in

280 f unctional or tissue injury. Can also the composed of the preceding elements of the cascade (classical/lectin pathway: C4bC2b complex, alternative pathway: C3bBb complex). Upon activation, the complement system can affect 281 activation, the complement system can affect cells in two ways: (3bBb complex). Upon activation, the complement system can affect cells in two ways: (1) by forming the membrane attack complex (MAC, sC5b-9 complex), resulti 282 (classical) only pathway. C20 Complex, alternative pathway: C20 Complex, and the membrane attack complex (MAC, sC5b-9 complex), resulting in osmotic lysis of the targeted cell, and (2) indirect opsonization through the dep 283 284 285 286

activation, the complex (MAC, sC5b-9 complex), resulting in osmotic lysis of the targeted cell, and (2) indirect opsonization through the deposition of C3b on the surface of microbes, which facilitates phagocytosis by immu attack complex of the deposition through the deposition of C3b on the surface of microbes, which
facilitates phagocytosis by immune cells.
CR1 acts as a negative complement regulator, reducing C3 activation and tissue
de facilitates phagocytosis by immune cells.
 CR1 acts as a negative complement regulator, reducing C3 activation and tissue

deposition, by processing and bounding immune complexes, which then facilitates their

transfer t *CR1* acts as a negative compledeposition, by processing and boundinariansfer to the liver or spleen where relassical and lectin pathways, *CR1* has de the formation of C3-convertase. In the a cleavage of active C3b fragm 287 CR1 acts as a negative complement regulator, reducing C3 activation and tissue

288 deposition, by processing and bounding immune complexes, which then facilitates their

289 transfer to the liver or spleen where macr 288 transfer to the liver or spleen where macrophages ingest and eliminate them. In both the
classical and lectin pathways, CRI has decay-activating activity, in that its binding C4b prevents
the formation of C3-convertase. I 289 classical and lectin pathways, *CR1* has decay-activating activity, in that its binding C4b prevents
the formation of C3-convertase. In the alternative pathway, *CR1* then acts as a cofactor for the
cleavage of active C3b 290 classical and lectin pathways, CR1 has decay-activating activity, in that its binding C+b prevents
the formation of C3-convertase. In the alternative pathway, CR1 then acts as a cofactor for the
cleavage of active C3b frag 291 the formation of C3-convertase. In the alternative pathway, CAT then acts as a colactor for the
cleavage of active C3b fragments, which could activate C3.²¹ CR1 significantly reduces C3b
deposition by ~80% over the class 292 deposition of C3b fragments, which could activate C3.²¹ CR1 significantly reduces C3b
deposition by ~80% over the classical pathway, but *CR1* is most potent when the alternative
pathway is activated (> 95% reduction in deposition of C3b fragments, which could activate C3.²¹ CR1 significantly reduces C3b
deposition by ~80% over the classical pathway, but *CR1* is most potent when the alternative
pathway is activated (> 95% reduction in 294 pathway is activated (> 95% reduction in C3b deposition).²¹ By stopping the activation of the complement system at the stage of C3-convertase formation, *CR1* is also indirectly involved in reducing the deposition of sC5 296 297

296 pathway is activated (> 95% reduction in C3b deposition).²¹ By stopping the activation of the

296 complement system at the stage of C3-convertase formation, *CR1* is also indirectly involved in

297 reducing the de reducing the deposition of sCSb-9, which would have formed afterwards if the complement
system had been subsequently activated.
CR1 is expressed by several cell types, including red blood cells (RBC), leukocytes, and
amo EXECT IN A System had been subsequently activated.

CR1 is expressed by several cell types, including red blood cells (RBC), leukocytes, and

among specialized renal cells, CR1 is localized exclusively on glomerular podoc 298 system had been subsequently activated.

299 $CR1$ is expressed by several cell t

300 among specialized renal cells, $CR1$ is loc.

201 related variants in $CR1 - rs17047661$

302 individuals and its pair – $rs17047660$ are *CR1* is expressed by several cell types, including red blood cells (RBC), leukocytes, and

300 among specialized renal cells, *CR1* is localized exclusively on glomerular podocytes. The FSGS

301 related variants in *CR1* 300 among specialized renal cells, CR1 is localized exclusively on glomerular podocytes. The F5GS
related variants in CR1 – rs17047660 are known as Knops group polymorphisms and are a part
of the Red Cell Surface Antigens, wh 301 related variants in CR1 – 1317047661 which was Tepheated in Latin-American descent
individuals and its pair – rs17047660 are known as Knops group polymorphisms and are a part
of the Red Cell Surface Antigens, which give r 302 individuals and its pair of the Red Cell Surface Antigens, which give rise to the Sl2 and McC^b alleles in the Swain-Langley 1 and 2 allele pairs (Sl1/Sl2) and McCoy a and b (McC^a/McC^b), respectively. ²². The K1590 of the Red Cell Surface Antigens, which give rise to the Sl2 and McC^b alleles in the Swain-Langley
304 1 and 2 allele pairs (Sl1/Sl2) and McCoy a and b (McC^a/McC^b), respectively. ²². The K1590E
305 substitution an 1 and 2 allele pairs (Sl1/Sl2) and McCoy a and b (McC^a/McC^b), respectively. ²². The K1590E
305. substitution and the R1601G substitution in *CR1* are located just 11 amino acids away from
306. each other and are in 305 substitution and the R1601G substitution in CR1 are located just 11 amino acids away from
each other and are in strong LD for both the 1000 genomes for the global African population
12 306

It is made available under a CC-BY 4.0 International license.

308

(r²=0.15, D'=1) and for African-American descent from study data (r²=0.24, D'=1). They are

308 located in the Long Homologous Repeats (LHRs), motif D, which is responsible for C1q and MBL

309 binding.²²

310 No ge Mo genetic alterations in the adaptive immune system have been identified in FSGS
patients to date, and the role of immune and complement systems genetic variants remains
unknown. Recent studies have described the role of 309 binding. ²²
310 No
311 patients to
312 unknown.
313 glomerulo
314 disease, as
315 against kio
316 are causa
317 inflammat 310 No date, and the role of immune and complement systems genetic variants remains

2015 to date, and the role of immune and complement systems genetic variants remains

2016 the street studies have described the role of the 311 patition. Recent studies have described the role of the complement system in various
glomerulopathies.²³ Typically, the reduction in *CR1* expression is linked to the severity of the
disease, as indicated by the degree o 312 glomerulopathies.²³ Typically, the reduction in *CR1* expression is linked to the severity of the disease, as indicated by the degree of inflammation or tissue damage. ²⁴ Autoantibodies directed against kidney-expresse glomerulopathies.²³ Typically, the reduction in *CR1* expression is linked to the severity of the
disease, as indicated by the degree of inflammation or tissue damage.²⁴ Autoantibodies directed
against kidney-expresse disease, as indicated by the degree of inflammation or tissue damage. ²⁴ Autoantibodies directed
against kidney-expressed autoantigens or antibody/antigen complexes deposited in the kidney
are causative agents of variou 315 and causative agents of various human kidney diseases. There are cases of C3-mediated
inflammation and deposition. 25.26 Further, inhibition of C3 reduces proteinuria in animal
models.²⁷ With regard to FSGS, IgG and C3 d 316 inflammation and deposition. ^{25,26} Further, inhibition of C3 reduces proteinuria in animal
models.²⁷ With regard to FSGS, IgG and C3 deposits are often observed in the affected glomeruli,
but the pathogenic role of the 319

inflammation and deposition. ^{25,26} Further, inhibition of C3 reduces proteinuria in animal
318 models.²⁷ With regard to FSGS, lgG and C3 deposits are often observed in the affected glomeruli,
319 but the pathogenic rol models.²⁷ With regard to FSGS, IgG and C3 deposits are often observed in the affected glomeruli,
319 but the pathogenic role of these deposits remains still unclear, and therapy against the
320 complement system has not but the pathogen and urine of patients at
monstrated elevated levels of Ba, Bb, C4a, and sC5b-9 in the
plasma and urine of patients afflicted with primary FSGS. The detection of these protein
deposits in the blood signifie complement system has not been studied in FSGS.²⁸
321 Another study has demonstrated elevated
322 plasma and urine of patients afflicted with prin
323 deposits in the blood signifies that the compler
524 fragments can ga 321 and urine of patients afflicted with primary FSGS. The detection of these protein
s in the blood signifies that the complement cascade is activated at a site where
ts can gain entry into the vascular space, likely in the m 322 plasma and after the patition and the printing term and printing the detection of deposits in the blood signifies that the complement cascade is activated at a site where fragments can gain entry into the vascular space, l 323 ragments can gain entry into the vascular space, likely in the mesangial and sclerotic regions.
Conversely, the rise in urine Ba, C4a, and sC5b-9 levels in some patients may reflect
complement activation in the glomerulus, 324 for the risk in the value of the space of the variant space of the value of the value of the value of the complement activation in the glomerulus, or alternatively, activation of filtered proteins in the tubular lumen or d 325 Complement activation in the glomerulus, or alternatively, activation of filtered proteins in the tubular lumen or downstream in the urinary collection system. ²⁹ While C5b-9 complexes generally form directly on the memb 326 complement activation in the glomerating of anticiants of microsystem.²⁹ While C5b-9 complexes
generally form directly on the membranes of microorganisms, particularly Gram-negative ones,
they can also affect adjacent ce tubular lumen or downstream in the urinary collection system. ²⁹ While C5b-9 complexes
generally form directly on the membranes of microorganisms, particularly Gram-negative ones,
they can also affect adjacent cells, res 328 329

generally form directly on the memberally controlled in "bystander" harm.

The relationship between activation of the classical and alternative pathways in

response to the presence of Knops antigens in *CR1* has been inve The relationship between activation of the classical and
response to the presence of Knops antigens in CR1 has been inves
lack of correlation has been noted. However, there is a reasona
13 330 The relation has been antigens in CR1 has been investigated previously, and a
torrelation has been noted. However, there is a reasonable discrepancy between
13 331 response to the presence of Knops antigens in CR1 has been investigated previously, and a
lack of correlation has been noted. However, there is a reasonable discrepancy between
13 332 $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 &$

It is made available under a CC-BY 4.0 International license.

333 the results of series of Shifted Terminal studies of the conjecture that SI2 and
McC^b influence the phenotype by modulating the activity of the cofactor implicated in the
cleavage of C3b and C4b or the C1q binding activi 334 Fridance the phenotype by modulating the activity of the cofactor implicated in the cleavage of C3b and C4b or the C1q binding activity. ³⁰ Nevertheless, these findings warrant future exploration of the involvement of t McC^b influence the phenotype by modulating the activity of the cofactor implicated in the

deavage of C3b and C4b or the C1q binding activity.³⁰ Nevertheless, these findings

warrant future exploration of the involvem Glassaure of C3b and C4b or the C1q binding activity.³⁰ Nevertheless, these findings
337 warrant future exploration of the involvement of the lectin pathway in the activation of
338 the complement system.³¹ For exampl 337 Frantistical of the complement system. ³¹ For example, the contribution of the lectin pathway in the activation of the complement system may play a crucial role in the development of progressive glomerular damage and lon the complement system. ³¹ For example, the contribution of the lectin pathway in the activation of the complement system may play a crucial role in the development of progressive glomerular damage and long-term urinary 339 progressive glomerular damage and long-term urinary abnormalities in patients with
Henoch-Schönlein purpura nephritis (HSPN).³² In the case of FSGS, the presence of MBL
deposits that are focal and segmental has been obs 340 From Section Exhibition purpura nephritis (HSPN).³² In the case of FSGS, the presence of MBL
deposits that are focal and segmental has been observed, as reported in previous studies,
and this can also result in tissue d 342 343

Henoch-Schönlein purpura nephritis (HSPN). ³² In the case of FSGS, the presence of MBL
deposits that are focal and segmental has been observed, as reported in previous studies,
343 and this can also result in tissue dam and this can also result in tissue damage, MBL deficiency can lead to autoimmune
diseases. 33,34
There are many different pathogens that use *CR1* as a receptor for cell entry
example: *Leishmania major* ∞ , *Legionell* diseases. 33,34
There are many different pathogens that use *CR1* as a receptor for cell entry
example: *Leishmania major* ³⁵, *Legionella pneumophila* ³⁶, *Leishmania panamensis* ³⁷
Mycobacterium tuberculosis ³⁸ 344 diseases. ^{33,34}
345 There
346 example: *Le*
347 Mycobacteriu
348 Plasmodium_J
349 hypothesis, :
350 malaria in su
351 African-desce
352 was noted by 345 External many different pathogens that use CR1 as a receptor for each entry, for

Exterium tuberculosis ³⁸. It also has been shown that CR1 is a RBC receptor used by
 dium falciparum for cell invasion, independent of example: Leishmania major ³⁵, Legionella pneumophila ³⁶, Leishmania panamensis ³⁷ and

347 *Mycobacterium tuberculosis* ³⁸. It also has been shown that *CR1* is a RBC receptor used by

348 *Plasmodium falciparum* 347 Mycobacterium tuberculosis 38 . It also has been shown that CR1 is a RBC receptor used by . It also has been shown that ch1 is a RBC receptor used by
all invasion, independent of sialic acid.³⁹ Consistently with this
a NBC receptor used by
a provided in much higher prevalence of SI2 in
ompared to Europeans. *Plasmodium falciparum* for cell invasion, independent of sialic acid.³⁹ Consistently with this

349 hypothesis, Sl2 (rs17047661) was previously associated with protection against cerebral

350 malaria in sub-Saharan Af 349 malaria in sub-Saharan African populations, which resulted in much higher prevalence of SI2 in
African-descent individuals compared to Europeans. ¹⁹ In the study conducted by Opi et al, it
was noted by the authors that t 350 African-descent individuals compared to Europeans.¹⁹ In the study conducted by Opi et al, it
was noted by the authors that the opposite concomitant effect of the McC^b (rs17047660) allele
on the development of severe m African-descent individuals compared to Europeans. ¹⁹ In the study conducted by Opi et al, it

352 was noted by the authors that the opposite concomitant effect of the McC^b (rs17047660) allele

353 on the development was noted by the authors that the opposite concomitant effect of the McC^b (rs17047660) allele
on the development of severe malaria was of only nominal borderline significance, despite
being under significant strong posi 353 being under significant strong positive selection (iHS<-2.0)¹⁹. At first glance, the association with severe malaria and significant positive McC^b selection are discouraging, but this may explain the linkage of negativ being under significant strong positive selection (iHS<-2.0)¹⁹. At first glance, the association with severe malaria and significant positive McC^b selection are discouraging, but this may explain the linkage of negati with severe malaria and significant positive McC^b selection are discouraging, but this may

356 explain the linkage of negative and protection haplotypes to each other. Moreover, the authors

357 of the original article 356 explain the linkage of state allows and protocol of the original article tested some haplotypes of SI and McC combinations and found that the combination of SI2/McC^a alleles has an additive protective effect against mala 357 combination of Sl2/McC^a alleles has an additive protective effect against malaria, which may
explain the lack of replication signal for rs17047660 in the African-American descent.¹⁹
14 358 combination of SI2/McC^a alleles has an additive protective effect against malaria, which may
explain the lack of replication signal for rs17047660 in the African-American descent.¹⁹
14 explain the lack of replication signal for $rs17047660$ in the African-American descent.¹⁹
14

It is made available under a CC-BY 4.0 International license.

367

```
368 Data availability
```
- 369
- For the contribution and spin to a matrix and generative frequencies are available through
AP, accession numbers are available in the Supplementary Table 1.
contributions
I., J.B.K., C.A.W., M.J.D., A.S., M.A. designed a 370

mechanisms underlying the disease.
Data availability
FSGS cohort allele frequencie
Supplementary Tables. Raw sequence
the dbGAP, accession numbers are av
Author contributions Supplementary Table 1.

the dbGAP, accession numbers are available in the Supplementary Table 1.

Author contributions

R.S., Z.M., J.B.K., C.A.W., M.J.D., A.S., M.A. designed and conceived the study.

R.S., Z.M., A.L., J. 371

372

373 Author contributions

- Author contributions
R.S., Z.M., J.B.K., C.A.W., M.J.D., A.S., M.A. designed and conceived the study.
R.S., Z.M., A.L., J.B.K., C.A.W., Al.S., A.S., M.J.D., M.A. analyzed the data
J.B.K., C.A.W., M.J.D., A.S., M.A. acquire 374
- R.S., Z.M., A.L., J.B.K., C.A.W., Al.S., A.S., M.J.D., M.A. analyzed the data
J.B.K., C.A.W., M.J.D., A.S., M.A. acquired funding
N.G., C.S. managed control cohorts
M.J.D., A.S., M.A. supervised the study
R.S., J.B.K., C.A 375
- 376
- 377
- M.J.D., A.S., M.A. supervised the stu
R.S., J.B.K., C.A.W., M.J.D., A.S., M.A. v
All authors reviewed and approved
Acknowledgments. This work w
NIDDK, NIH (JBK). This project ha 378
- N.G., C.S. managed control cohorts
M.J.D., A.S., M.A. supervised the study
R.S., J.B.K., C.A.W., M.J.D., A.S., M.A. wrote the ma
All authors reviewed and approved the manusc
Acknowledgments. This work was supported
NIDDK R.S., J.B.K., C.A.W., M.J.D., A.S., M.A. wro
All authors reviewed and approved the
Acknowledgments. This work was
NIDDK, NIH (JBK). This project has b 379
- 380

- R.B.K., C.A.W., M.J.D., A.S., M.A. acquired funding
N.G., C.S. managed control cohorts
M.J.D., A.S., M.A. supervised the study
R.S., J.B.K., C.A.W., M.J.D., A.S., M.A. wrote the manuscript
All authors reviewed and approved All authors reviewed and approved the manuscript
Acknowledgments. This work was supported in part
NIDDK, NIH (JBK). This project has been funded in part
15 Acknowledgments. This work was supported in
NIDDK, NIH (JBK). This project has been funded in
- 382 Acknowledgments. This work was supported in part by the Intramural Research Program,
383 NIDDK, NIH (JBK). This project has been funded in part with federal funds from the National
15 383 $\frac{1}{2}$ and $\frac{1}{2}$. This project has been fund in part with federal funds from the National fund

It is made available under a CC-BY 4.0 International license.

384 The manning matrix in the manning term in the contract 7500 clicles of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Gover 385 THET INTERT INTERT INTERT INTERTM INTERTM INTERTM INTERTMENT INTERTMENT INTERTMENT INTERTMENT IN the U.S. Government. This Research was supported [in part] by the Intramural Research Program of the NIH, National Cancer Ins 386 and Human Services, and the Minister mention in Figures mention products, to regalization
imply endorsement by the U.S. Government. This Research was supported [in part] by the
Intramural Research Program of the NIH, Natio 387 inframural Research Program of the NIH, National Cancer Institute, Center for Cancer Research
and NIDDK. The authors acknowledge the contributions of the following investigators who
recruited subjects for FSGS genetic stud 388 and NIDDK. The authors acknowledge the contributions of the following investigators who
recruited subjects for FSGS genetic studies, published in Kopp et al, Nature Genetics, 2008, as
DNA from those subjects was used in th 389 and MIDDK and MIDDK. The authority of Science in Mind Mind of the Russian DNA from those subjects was used in the present study. These investigators include Kopp JB, Freedman BI, Ahuja TS, Berns JS, Briggs W, Cho ME, Dart 390 recruits for the subjects was used in the present study. These investigators include Kopp JB,
Freedman BI, Ahuja TS, Berns JS, Briggs W, Cho ME, Dart RA, Kimmel PL, Korbet SM, Michel DM,
Mokrzycki MH, Schelling JR, Simon E 391 Freedman BI, Ahuja TS, Berns JS, Briggs W, Cho ME, Dart RA, Kimmel PL, Korbet SM, Michel DM,
Mokrzycki MH, Schelling JR, Simon E, Trachtman H.
R.S., Al.S. were supported by the Ministry of Science and Higher Education of t 392 393

- 394 395
- Freedman BI, Energy Press, Britan BI, The Free, Education Captures Journal Pression
Mokrzycki MH, Schelling JR, Simon E, Trachtman H.
R.S., Al.S. were supported by the Ministry of Science and Higher Education of the Russia R.S., Al.S. were supported by the Ministry of Science Federation (Priority 2030 Federal Academic Leader
A.L. was supported by Ministry of Science and H
(Agreement # 075-15-2022-301).
M.A. was in part supported by Nationwid Federation (Priority 2030 Federal Academic Leadership Program).
A.L. was supported by Ministry of Science and Higher Education of the Russian Federation
(Agreement # 075-15-2022-301).
M.A. was in part supported by Nationwi Federal A.L. was supported by Ministry of Science and Higher Education
(Agreement # 075-15-2022-301).
M.A. was in part supported by Nationwide Foundation Pediatric Inn
Disclosures
M.J.D. is a founder of Maze Therapeutics; 396 A. L. was in part supported by Nationwide Foundation Pediatric Innovation Fund.

M.A. was in part supported by Nationwide Foundation Pediatric Innovation Fund.

Disclosures

M.J.D. is a founder of Maze Therapeutics; A.S. a 397
- M.A. was in part supported by Nati
Disclosures
M.J.D. is a founder of Maze Thera
Other authors have no competing 398

399

400 Disclosures

- Disclosures
M.J.D. is a founder of Maze Therapeutics; A.S. and Z.M. are employees of Genen
Other authors have no competing interests to disclose. 401
- M.J.D. is a f ounder of Maze Therapeutics; A.S. and Z.M. are employees of Generations
Other authors have no competing interests to disclose. 402 Other authors have no competing interests to disclose.

It is made available under a CC-BY 4.0 International license.

403 References

- 405
- 406
- 407
- 408
- 409 410
- 411
- 412
- (2018). Glomerular disease frequencies by race, sex and region: results from the Internation
Kidney Biopsy Survey. Nephrol. Dial. Transplant. 33, 661–669.
2. Dragovic, D., Rosenstock, J.L., Wahl, S.J., Panagopoulos, G., De Kidney Biopsy Survey. Nephrol. Dial. Transplant. 33, 661–669.
2. Dragovic, D., Rosenstock, J.L., Wahl, S.J., Panagopoulos, G., DeVita, M.V., and Michelis, M.F.
(2005). Increasing incidence of focal segmental glomeruloscler Kidney Biopsy Survey. Rephrol. Dial. Transplant. 35, 661–669.
2. Dragovic, D., Rosenstock, J.L., Wahl, S.J., Panagopoulos, G., De
(2005). Increasing incidence of focal segmental glomeruloscler
demographic patterns. Clin. N (2005). Increasing incidence of focal segmental glomerulosclerosis and an examination of demographic patterns. Clin. Nephrol. 63, 1–7.
3. Shabaka, A., Tato Ribera, A., and Fernández-Juárez, G. (2020). Focal Segmental
Glome demographic patterns. Clin. Nephrol. 63, 1–7.
3. Shabaka, A., Tato Ribera, A., and Fernández-Juárez, G. (2020). Focal Segmental
Glomerulosclerosis: State-of-the-Art and Clinical Perspective. Nephron 144, 413–427.
4. Kitiya demographic patterns. Clin. Nephrol. 65, 1–7.
3. Shabaka, A., Tato Ribera, A., and Fernández-
Glomerulosclerosis: State-of-the-Art and Clini
4. Kitiyakara, C., Kopp, J.B., and Eggers, P. (200
glomerulosclerosis. Semin. Nep Glomerulosclerosis: State-of-the-Art and Clinical Perspective. Nephron 144, 413-4. Kitiyakara, C., Kopp, J.B., and Eggers, P. (2003). Trends in the epidemiology of f
glomerulosclerosis. Semin. Nephrol. 23, 172–182.
5. Yu, 413 414
- 415
- 416 417
-
- 418
- 419
- 420
- 421
- 422 423
- d. Kitiyakara, C., Kopp, J.B., State-of-the-Art and chinden 1 erspective. Nephron 144, 413–427.
4. Kitiyakara, C., Kopp, J.B., and Eggers, P. (2003). Trends in the epidemiology of focal s
glomerulosclerosis. Semin. Nephrol 1. Sponsor Unity, H., Kitiyama, H., Abuja, T.S., et al. (2008). MYH9 is a major-effect risk gene for focal segmental of experimental and F. (2016). A role for genetic susceptibility in sporadic focal segmental glomeruloscl glomer doseterosis. Semin. Nephrol. 25, 172–162.
5. Yu, H., Artomov, M., Brähler, S., Stander, M.C., Sh
M., Miner, J.H., Jain, S., et al. (2016). A role for gene
glomerulosclerosis. J. Clin. Invest. 126, 1067–1078
6. Kopp, M., Miner, J.H., Jain, S., et al. (2016). A role for genetic susceptibility in sporadic focal segmental
glomerulosclerosis. J. Clin. Invest. 126, 1067-1078.
6. Kopp, J.B., Nelson, G.W., Sampath, K., Johnson, R.C., Genovese glomerulosclerosis. J. Clin. Invest. 126, 1067–1078.
6. Kopp, J.B., Nelson, G.W., Sampath, K., Johnson, R.C., Genovese, G., An, P., Friedman, D., Briggs, W., Dart, R., Korbet, S., et al. (2011). APOL1 genetic variants in f glomer diosclerosis. J. Clin. Invest. 126, 1007–1070.
6. Kopp, J.B., Nelson, G.W., Sampath, K., Johnson, R.C
W., Dart, R., Korbet, S., et al. (2011). APOL1 genetic
glomerulosclerosis and HIV-associated nephropath
7. Kopp, W., Dart, R., Korbet, S., et al. (2011). APOL1 genetic variants in focal segmental
glomerulosclerosis and HIV-associated nephropathy. J. Am. Soc. Nephrol. 22, 2129–2137.
7. Kopp, J.B., Smith, M.W., Nelson, G.W., Johnson, R glomerulosclerosis and HIV-associated nephropathy. J. Am. Soc. Nephrol. 22, 27. Kopp, J.B., Smith, M.W., Nelson, G.W., Johnson, R.C., Freedman, B.I., Bowden, I
McKenzie, L.M., Kajiyama, H., Ahuja, T.S., et al. (2008). MYH9 glomer abscretosis and HIV-associated nephropathy. J. Am. Soc. Nephrol. 22, 2129–2137.
7. Kopp, J.B., Smith, M.W., Nelson, G.W., Johnson, R.C., Freedman, B.I., Bowden, D.W., Oleks;
McKenzie, L.M., Kajiyama, H., Ahuja, T.S. McKenzie, L.M., Kajiyama, H., Ahuja, T.S., et al. (2008). MYH9 is a major-effect risk gene for focal
segmental glomerulosclerosis. Nat. Genet. 40, 1175–1184.
8. Kao, W.H.L., Klag, M.J., Meoni, L.A., Reich, D., Berthier-Sch esgmental glomerulosclerosis. Nat. Genet. 40, 1175–1184.

8. Kao, W.H.L., Klag, M.J., Meoni, L.A., Reich, D., Berthier-Schaad, Y., Li, M., Coresh, J., Patterson, N.

Tandon, A., Powe, N.R., et al. (2008). MYH9 is associate segmental glomer diosclerosis. Nat. denet. 40, 1175–1164.
8. Kao, W.H.L., Klag, M.J., Meoni, L.A., Reich, D., Berthier-Schandon, A., Powe, N.R., et al. (2008). MYH9 is associated w
disease in African Americans. Nat. Genet. 424 425 426
- Tandon, A., Powe, N.R., et al. (2008). MYH9 is associated with nondiabetic end-stage renal

disease in African Americans. Nat. Genet. 40, 1185–1192.

9. Genovese, G., Friedman, D.J., Ross, M.D., Lecordier, L., Uzureau, P., disease in African Americans. Nat. Genet. 40, 1185–1192.
9. Genovese, G., Friedman, D.J., Ross, M.D., Lecordier, L., Uzureau, P., Freedman, B.I., Bowde
D.W., Langefeld, C.D., Oleksyk, T.K., Uscinski Knob, A.L., et al. (201 9. Genovese, G., Friedman, D.J., Ross, M.D., Lecordier, L., U.
D.W., Langefeld, C.D., Oleksyk, T.K., Uscinski Knob, A.L., et
ApoL1 variants with kidney disease in African Americans
10. Van der Auwera, G.A., Carneiro, M.O., 427 428 429
- 430
- 431
- 432
-
- 434
- 435 436
- 19. U.N., Langefeld, C.D., Oleksyk, T.K., Uscinski Knob, A.L., et al. (2010). Association of trypanoly

19. W., Langefeld, C.D., Oleksyk, T.K., Uscinski Knob, A.L., et al. (2010). Association of trypanoly

ApoL1 variants w Apol.1 variants with kidney disease in African Americans. Science 329, 841–845.

10. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A.,

Jordan, T., Shakir, K., Roazen, D., Thib ApoL1 variants with Kidney disease in African Americans. Science 329, 041–045.

10. Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-M

Jordan, T., Shakir, K., Roazen, D., Thibault, J., et a Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence
variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics
11, 11.10.1-11.10.33.
11. Variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatic:
11, 11.10.1-11.10.33.
11. Hail Team. Hail 0.2.
12. Athey, T.L., Liu, T., Pedigo, B.D., and Vogelstein, J.T. (2019). AutoGMM: 11, 11, 10, 1-11, 10, 1-11, 10, 1-11, 10, 2.
11. Hail Team. Hail 0.2.
12. Athey, T.L., Liu, T., Pedigo, B.D., and Vogelstein, J.T. (2019). AutoGMM: Automatic and
Hierarchical Gaussian Mixture Modeling in Python. ArXiv.
13. 433 11, 11.10.1-11.10.33.
434 11. Hail Team. Hail 0.:
435 12. Athey, T.L., Liu, T.,
436 Hierarchical Gaussian
437 13. Ho, D.E., Imai, K., I
60 Ferametric Causal
439 14. Wang, M., Chun, J.,
440 G.B., Lifton, R.P., Mane
441 12. Athey, T.L., Liu, T., P
Hierarchical Gaussian M
13. Ho, D.E., Imai, K., Kir
for Parametric Causal Ir
14. Wang, M., Chun, J., G
G.B., Lifton, R.P., Mane, S
sporadic FSGS. J. Am. So
15. Edmonson, M.N., Pa Hierarchical Gaussian Mixture Modeling in Python. ArXiv.
13. Ho, D.E., Imai, K., King, G., and Stuart, E.A. (2011). Matchlt@: Nonparametric Preproce:
for Parametric Causal Inference. J. Stat. Softw. 42,.
14. Wang, M., Chun 13. Ho, D.E., Imai, K., King, G., and Stuart, E.A. (2011). Matc
for Parametric Causal Inference. J. Stat. Softw. 42,.
14. Wang, M., Chun, J., Genovese, G., Knob, A.U., Benjamin, .
G.B., Lifton, R.P., Mane, S., et al. (2019 437 13. Homewitch Causal Inference. J. Stat. Softw. 42,
14. Wang, M., Chun, J., Genovese, G., Knob, A.U., Benjamin, A., Wilkins, M.S., Friedman, D.J., Appe
G.B., Lifton, R.P., Mane, S., et al. (2019). Contributions of rare gen 438
- 439
- for Farametric Causal Inference. J. Stat. Softw. 42,.
14. Wang, M., Chun, J., Genovese, G., Knob, A.U., Bel
G.B., Lifton, R.P., Mane, S., et al. (2019). Contributic
sporadic FSGS. J. Am. Soc. Nephrol. *30*, 1625–1640
15. E 14. G.B., Lifton, R.P., Mane, S., et al. (2019). Contributions of rare gene variants to familial and
sporadic FSGS. J. Am. Soc. Nephrol. 30, 1625–1640.
15. Edmonson, M.N., Patel, A.N., Hedges, D.J., Wang, Z., Rampersaud, E 440
- 441
- sporadic FSGS. J. Am. Soc. Nephrol. *30*, 1625–1640.
15. Edmonson, M.N., Patel, A.N., Hedges, D.J., Wang, Z., Rampersaud, E., Kesserwan, C.A., Zh.
17 sporadic FSGS. J. Am. Soc. Nephrol. 30, 1023–1640.
15. Edmonson, M.N., Patel, A.N., Hedges, D.J., Wang,
17 442 $17.$ Edmonstration, $\frac{1}{2}$. $\frac{1}{2}$.

443 Information Exchange (PeCanPIE): a cloud-based platform for curating and classifyir
variants. Genome Res. 29, 1555–1565.
16. Chennen, K., Weber, T., Lornage, X., Kress, A., Böhm, J., Thompson, J., Laporte, J., a
(2020). MI 444 Variants. Genome Res. 29, 1555–1565.

16. Chennen, K., Weber, T., Lornage, X., Kress, A., Böhm, J., Thompson, J., Laporte, J., and Poch, O.

(2020). MISTIC: A prediction tool to reveal disease-relevant deleterious missens 445 variants. Genome Res. 29, 1555–1565.
446 16. Chennen, K., Weber, T., Lornage, X.,
447 (2020). MISTIC: A prediction tool to re
0NE 15, e0236962.
449 17. Samocha, K.E., Kosmicki, J.A., Karcz
450 MacArthur, D.G., Neale, B 446 (2020). MISTIC: A prediction tool to reveal disease-relevant deleterious missense variants. PLoS
ONE 15, e0236962.
17. Samocha, K.E., Kosmicki, J.A., Karczewski, K.J., O'Donnell-Luria, A.H., Pierce-Hoffman, E.,
MacArthur, 447 ONE 15, e0236962.
17. Samocha, K.E., Kosmicki, J.A., Karczewski, K.J., O'Donnell-Luria, A.H., Pierce-Hoffman, E.,
MacArthur, D.G., Neale, B.M., and Daly, M.J. (2017). Regional missense constraint improves
variant deleteri 448 ONE 15, e0236962.

449 17. Samocha, K.E., K

450 MacArthur, D.G., Ne

451 variant deleteriousr

452 18. Lee, S., Abecasis,

53 study designs and s

454 19. Opi, D.H., Swann

455 Kone, A.K., Diallo, D.

456 associations 449 MacArthur, D.G., Neale, B.M., and Daly, M.J. (2017). Regional missense constraint improves
variant deleteriousness prediction. BioRxiv.
18. Lee, S., Abecasis, G.R., Boehnke, M., and Lin, X. (2014). Rare-variant association 450 variant deleteriousness prediction. BioRxiv.

18. Lee, S., Abecasis, G.R., Boehnke, M., and Lin, X. (2014). Rare-variant association analysis

study designs and statistical tests. Am. J. Hum. Genet. 95, 5–23.

19. Opi, D. 451 18. Lee, S., Abecasis, G.R., Boehnke, M., and L
study designs and statistical tests. Am. J. Hur
19. Opi, D.H., Swann, O., Macharia, A., Uyoga,
Kone, A.K., Diallo, D.A., et al. (2018). Two con
associations with cerebral mal 452 study designs and statistical tests. Am. J. Hum. Genet. 95, 5–23.
19. Opi, D.H., Swann, O., Macharia, A., Uyoga, S., Band, G., Ndila, C.M., Harrison, E.M., Thera, M.
Kone, A.K., Diallo, D.A., et al. (2018). Two complement 453 study designs and statistical tests. Am. J. Hum. denet. *55*, 5–25.
19. Opi, D.H., Swann, O., Macharia, A., Uyoga, S., Band, G., Ndila,
Kone, A.K., Diallo, D.A., et al. (2018). Two complement receptor
associations with ce 454 Kone, A.K., Diallo, D.A., et al. (2018). Two complement receptor one alleles have opposing
associations with cerebral malaria and interact with α+thalassaemia. ELife 7,
20. Freiwald, T., and Afzali, B. (2021). Renal disea 455 associations with cerebral malaria and interact with α +thalassaemia. ELife 7,
20. Freiwald, T., and Afzali, B. (2021). Renal diseases and the role of complement: Linking
complement to immune effector pathways and therap 456 associations with cerebral malaria and interact with α+thalassaemia. Elife 7,.
20. Freiwald, T., and Afzali, B. (2021). Renal diseases and the role of complement
complement to immune effector pathways and therapeutics. Ad 457 complement to immune effector pathways and therapeutics. Adv. Immunol. 152, 1–81.
21. Poppelaars, F., and Thurman, J.M. (2020). Complement-mediated kidney diseases. Mol.
Immunol. 128, 175–187.
22. Moulds, J.M. (2010). The 458 complement to immune enector pathways and therapeutes. Adv. immunol. 152, 1–01.
21. Poppelaars, F., and Thurman, J.M. (2020). Complement-mediated kidney diseases. M
munol. 128, 175–187.
22. Moulds, J.M. (2010). The Knops b 459 12. Moulds, J.M. (2010). The Knops blood-group system: a review. Immunohematology 26,
22. Moulds, J.M. (2010). The Knops blood-group system: a review. Immunohematology 26,
23. Mathern, D.R., and Heeger, P.S. (2015). Molecu 460 mmunol. 128, 175–187.
22. Moulds, J.M. (2010). 1
23. Mathern, D.R., and He
Clin. J. Am. Soc. Nephrol.
24. Moll, S., Miot, S., Sada
complement receptor 1 s
168.
25. Willows, J., Wood, K.,
presenting with nephroti
26. Sethi, 461 22. Moulds, J.M. (2010). The Kings blood-group system: a review. Immunohematology 26, 2–7.
23. Mathern, D.R., and Heeger, P.S. (2015). Molecules great and small: the complement system.
24. Moll, S., Miot, S., Sadallah, S., 462 Clin. J. Am. Soc. Nephrol. 10, 1636–1650.

24. Moll, S., Miot, S., Sadallah, S., Gudat, F., Mihatsch, M.J., and Schifferli, J.A. (2001). No

complement receptor 1 stumps on podocytes in human glomerulopathies. Kidney Int. 463 cm. J. Am. soc. Nephrol. 10, 1030–1630.
24. Moll, S., Miot, S., Sadallah, S., Gudat, F.
complement receptor 1 stumps on podoc
168.
25. Willows, J., Wood, K., Bourne, H., and
presenting with nephrotic syndrome. BM
26. Sethi 464 complement receptor 1 stumps on podocytes in human glomerulopathies. Kidney Int.
168.
25. Willows, J., Wood, K., Bourne, H., and Sayer, J.A. (2019). Acquired C1-inhibitor defic
presenting with nephrotic syndrome. BMJ Case 465 compenent receptor 1 stumps on podocytes in numan glomer diopadnes. Kidney int. 59, 100–25. Willows, J., Wood, K., Bourne, H., and Sayer, J.A. (2019). Acquired C1-inhibitor deficiency presenting with nephrotic syndrome. BM 466 25. V
prese
26. S
Smitlabno
27. S
expe
28. S
Thur
393–
29. T
M. K., Segm 467 presenting with nephrotic syndrome. BMJ Case Rep. 12,
26. Sethi, S., Fervenza, F.C., Zhang, Y., Zand, L., Vrana, J.A., Nasr, S.H., Theis, J.D., Dogan, A., and
Smith, R.J.H. (2012). C3 glomerulonephritis: clinicopathologica 468 presenting with nephrotic syndrome. BMJ case Rep. 12,.
26. Sethi, S., Fervenza, F.C., Zhang, Y., Zand, L., Vrana, J.A.,
Smith, R.J.H. (2012). C3 glomerulonephritis: clinicopatho
abnormalities, glomerular proteomic profile, 469 Sethi, R.J.H. (2012). Calcomer (2012). Can also the complement than the summary state and follow-up. Ridney Int. 82, 465–4
27. Salant, D.J., Belok, S., Madaio, M.P., and Couser, W.G. (1980). A new role for complement in
ex 470 abnormalities, glomerular proteomic profile, treatment, and follow-up. Kidney Int. 82,
27. Salant, D.J., Belok, S., Madaio, M.P., and Couser, W.G. (1980). A new role for compler
experimental membranous nephropathy in rats. 471 abnormalities, glomerular proteomic prome, treatment, and follow-up. Ndiffer 02, 465–473.
27. Salant, D.J., Belok, S., Madaio, M.P., and Couser, W.G. (1980). A new role for complement in
experimental membranous nephropathy 472 experimental membranous nephropathy in rats. J. Clin. Invest. 66, 1339–1350.
28. Strassheim, D., Renner, B., Panzer, S., Fuquay, R., Kulik, L., Ljubanović, D., Holers, V.M., and
Thurman, J.M. (2013). IgM contributes to glo 473 experimental membranous nephropathy in rats. J. Clin. Invest. 66, 1339–1350.
28. Strassheim, D., Renner, B., Panzer, S., Fuquay, R., Kulik, L., Ljubanović, D., Hc
Thurman, J.M. (2013). IgM contributes to glomerular injury 474 Thurman, J.M. (2013). IgM contributes to glomerular injury in FSGS. J. Am. Soc. Nephrol. 24, 393–406.
29. Thurman, J.M., Wong, M., Renner, B., Frazer-Abel, A., Giclas, P.C., Joy, M.S., Jalal, D., Radeva, M.K., Gassman, J., 475 Thurman, J.M. (2015). Ignt contributes to glomerular injury in F505. J. Am. Soc. Nephrol. 24, 0.93–406.
29. Thurman, J.M., Wong, M., Renner, B., Frazer-Abel, A., Giclas, P.C., Joy, M.S., Jalal, D., Radev.
M.K., Gassman, J. 476 29. Thurn
M.K., Gass
Segmenta
30. Tetteh
Atkinson,
of soluble
type 1 are 477 29. Thurman, J., Gipson, D.S., et al. (2015). Complement Activation in Patients with Focal Segmental Glomerulosclerosis. PLoS ONE 10, e0136558.
29. Tetteh-Quarcoo, P.B., Schmidt, C.Q., Tham, W.-H., Hauhart, R., Mertens, H. 478 M.K., Merten Comparental Glomerulosclerosis. PLoS ONE 10, e0136558.

30. Tetteh-Quarcoo, P.B., Schmidt, C.Q., Tham, W.-H., Hauhart, R., Mertens, H.D.T., Rowe, A.

Atkinson, J.P., Cowman, A.F., Rowe, J.A., and Barlow, P.N. 479 Segmental Glomerulosclerosis. I Eos ONE 10, e0136358.
30. Tetteh-Quarcoo, P.B., Schmidt, C.Q., Tham, W.-H., Hau
Atkinson, J.P., Cowman, A.F., Rowe, J.A., and Barlow, P.N.
of soluble protein fragments that Knops blood group 480 30. The Microsoftermannia, P.C., P. Hault, W.C., P. Hauth, W.C., P. (2012). Lack of evidence from student soluble protein fragments that Knops blood group polymorphisms in complement recepty 1 are driven by malaria. PLoS O 481 At the protein fragments that Knops blood group polymorphisms in complement receptor-
type 1 are driven by malaria. PLoS ONE 7, e34820.
 18 482 483 type 1 are driven by malaria. Thus ONE 7, e34820.

It is made available under a CC-BY 4.0 International license.

- (2000). Complement receptor 1/CD35 is a receptor for mannan-binding lectin. J. Exp. Med. 192,
1797–1808.
32. Roos, A., Rastaldi, M.P., Calvaresi, N., Oortwijn, B.D., Schlagwein, N., van Gijlswijk-Janssen,
D.J., Stahl, G.L. 485 486
- 487
- 488
- 489 490
- 491 492
- (2000). Complement receptor 1/CD35 is a receptor for mannian-binding lectin. J. Exp. Med. 122, 1797-1808.
32. Roos, A., Rastaldi, M.P., Calvaresi, N., Oortwijn, B.D., Schlagwein, N., van Gijlswijk-Janssen, D.J., Stahl, G.L 32. Roos, A.,
D.J., Stahl, G.
the lectin pat
disease. J. Ar.
33. Lhotta, K
1808. Tsutsum
autoimmune
35. Da Silva,
binding of in
Immunol. 14
36. Payne, N. disease. J. Am. Soc. Nephrol. 17, 1724–1734.

33. Lhotta, K., Würzner, R., and König, P. (1999). Glomerular deposition of mannose-binding

lectin in human glomerulonephritis. Nephrol. Dial. Transplant. 14, 881–886.

34. Ts disease. J. Am. 30c. Nephrol. 17, 1724–1734.
33. Lhotta, K., Würzner, R., and König, P. (194
lectin in human glomerulonephritis. Nephrol
34. Tsutsumi, A., Takahashi, R., and Sumida,
autoimmune disease. Autoimmun. Rev. 4, 3 lectin in human glomerulonephritis. Nephrol. Dial. Transplant. 14, 881–886.
34. Tsutsumi, A., Takahashi, R., and Sumida, T. (2005). Mannose binding lectin: genetics and
autoimmune disease. Autoimmun. Rev. 4, 364–372.
35. D lectin in human giomer ulonephritis. Wephrol. Dial. Transplant. 14, 001-000.
34. Tsutsumi, A., Takahashi, R., and Sumida, T. (2005). Mannose binding lecti
autoimmune disease. Autoimmun. Rev. 4, 364–372.
35. Da Silva, R.P., 493 494
- 495
- 496 497
- 12. B., Stahl, G.L., Matsushita, M., Fujita, T., van Kooten, C., et al. (2006). Glomerular activation of
the lectin pathway of complement in IgA nephropathy is associated with more severe renal
disease. J. Am. Soc. Nephrol the lectin pathway of complement in IgA nephropathy is associated with more severe renal
disease. J. Am. Soc. Nephrol. 17, 1724–1734.
33. Lhotta, K., Würzner, R., and König. P. (1999). Glomerular deposition of mannose-bind 24. The Cassac, Autoimmun Rev. 4, 364–372.

35. Da Silva, R.P., Hall, B.F., Joiner, K.A., and Sacks, D.L. (1989). CR1, the C3b receptor, mediat

binding of infective Leishmania major metacyclic promastigotes to human macro autommune uisease. Autommun. Rev. 4, 364–372.
35. Da Silva, R.P., Hall, B.F., Joiner, K.A., and Sacks, D.
binding of infective Leishmania major metacyclic pr
Immunol. 143, 617–622.
36. Payne, N.R., and Horwitz, M.A. (1987) 498 499
- 500
- 501
- 502
- 503
- 504 505
- 506
- binding of infective Leishmania major metacyclic promastigotes to human macrophages.].
1mmunol. 143, 617–622.
36. Payne, N.R., and Horwitz, M.A. (1987). Phagocytosis of Legionella pneumophila is mediated
by human monocyte Immunol. 143, 617–622.

36. Payne, N.R., and Horwitz, M.A. (1987). Phagocytosis of Legionella pneumophila is media

by human monocyte complement receptors. J. Exp. Med. 166, 1377–1389.

37. Robledo, S., Wozencraft, A., Val mmunol. 143, 617–622.
36. Payne, N.R., and Horv
by human monocyte com
37. Robledo, S., Wozencra
by Leishmania (Viannia)
susceptibility in vitro wit
38. Schlesinger, L.S., Belli
Phagocytosis of Mycobac
receptors and complem by human monocyte complement receptors, J. Exp. Med. 166, 1377–1389.
37. Robledo, S., Wozencraft, A., Valencia, A.Z., and Saravia, N. (1994). Human monocyte infection
by Leishmania (Viannia) panamensis. Role of complement by human monocyte complement receptors. J. Exp. med. 160, 1377–1389.
37. Robledo, S., Wozencraft, A., Valencia, A.Z., and Saravia, N. (1994). Humaby Leishmania (Viannia) panamensis. Role of complement receptors and c
susce by Leishmania (Viannia) panamensis. Role of complement receptors and correlation of
susceptibility in vitro with clinical phenotype. J. Immunol. 152, 1265–1276.
38. Schlesinger, L.S., Bellinger-Kawahara, C.G., Payne, N.R., sysceptibility in vitro with clinical phenotype. J. Immunol. 152, 1265–1276.

38. Schlesinger, LS., Bellinger-Kawahara, C.G., Payne, N.R., and Horwitz, M.A. (1990).

Phagocytosis of Mycobacterium tuberculosis is mediated b susceptibility in vitro with clinical phenotype. J. Immunol. 152, 1265–1276.
38. Schlesinger, L.S., Bellinger-Kawahara, C.G., Payne, N.R., and Horwitz, M.A
Phagocytosis of Mycobacterium tuberculosis is mediated by human mo Phagocytosis of Mycobacterium tuberculosis is mediated by human monocyte complement complement component C3. J. Immunol. 144, 2771–2780.
39. Tham, W.-H., Wilson, D.W., Lopaticki, S., Schmidt, C.Q., Tetteh-Quarcoo, P.B., Ba receptors and complement component C3. J. Immunol. 144, 2771–2780.
39. Tham, W.-H., Wilson, D.W., Lopaticki, S., Schmidt, C.Q., Tetteh-Quarcoo, P.B., Barlow, P.N.
Richard, D., Corbin, J.E., Beeson, J.G., and Cowman, A.F. (receptors and complement component C3. J. Immunol. 144, 2771–2700.
39. Tham, W.-H., Wilson, D.W., Lopaticki, S., Schmidt, C.Q., Tetteh-Quarco
Richard, D., Corbin, J.E., Beeson, J.G., and Cowman, A.F. (2010). Complem
host e Exchard, D., Corbin, J.E., Beeson, J.G., and Cowman, A.F. (2010). Complement receptor 1 is the
host erythrocyte receptor for Plasmodium falciparum PfRh4 invasion ligand. Proc Natl Acad
USA 107, 17327–17332.
Figure Captions 507 Richard, D., Corpis and Control and Corpis and Corpis and Corpis and Corpis and Corpis and Corpis USA 107, 17327–17332.

Figure Captions

Figure 1. Study design, principal component analysis, and quantile-quantile plots to 508 USA 107, 17327–17332.

Figure Captions

Figure 1. Study design, principal component analysis, and quantile-quantile plots to

identify calibration of synonymous variants between case and control cohorts.

(A) Case-control 509
- 510
- 511 Figure Captions
- 512 Figure 1. Study design, principal component analysis, and quantile-quantile plots to
- 513 identify calibration of synonymous variants between case and control cohorts.
-
- 515
- USA 107, 17327–17332.
Figure Captions
Figure 1. Study design,
identify calibration of s
(A) Case-control study
examined for coding var.
genes and also was subje
(B) Principal componer 516
- 514 (A) Case-control study design. DNA samples from FSGS cases and controls were
examined for coding variants in a podocyte exome panel gene panel composed of 2
genes and also was subjected to whole exome analysis.
(B) Pri examined for coming variants in a point pair gener pair gener pair temperature in 2002
genes and also was subjected to whole exome analysis.
(B) Principal component analysis illustrates case-control matching in European-
d genes and also was subjected to which where also alternated to the principal component analysis.
illustrates case-conderived, African-derived and Latin-American-derived-p
genetic segregation of these three populations.
19
- 517 (B) Principal component analysis illustrates case-control matching in European-
518 derived, African-derived and Latin-American-derived-populations and demonstrat
519 genetic segregation of these three populations.
519 518
- derived and Latin-dependence and Latin-American-derived-populations and demonstrates and demonstrates and demonstrates and demonstrates and demonstrates and demonstrates are $\frac{1}{2}$ 519 genetic segregation of these three populations. These three populations is the population of the populations of the populations of the populations. The populations is the populations of the populations of the populations o

It is made available under a CC-BY 4.0 International license.

- 522
- 523
- 524
- GC (G) and the statistic against the expected median test statistic under the null hypothesis,
in which there is no association for each variant. This test identifies systemic biases and
significant associations. Here, mos 525
- 526
- 520 (C) Quantile-quantile (QQ)-plots for the association study of the common
521 synonymous variants with gnomAD population specific allele frequency \ge
522 These plots illustrate case-control matching quality for the E synonymous variants with gnomAD population specific allele frequency ≥0.01).
522 These plots illustrate case-control matching quality for the European-derived (left),
523 African-derived (middle), Latin-American-derived (r African-derived (middle), Latin-American-derived (right) populations. The test lamt
GC (genomic inflation factor) for genome-wide association studies (GWAS) compare
median test statistic against the expected median test st median test statistic against the expected median test statistic and the number of space of space and
significant associations. Here, most of the points fall along the diagonal, indicating the
absence of systemic bias.
Abb in which there is no associations. Here, most of the points fall along the diagonal, indicating the
absence of systemic bias.
Abbreviations. EUR, European-Americans. AFR, African-Americans. AMR, Admixture
Americans 527
- 528
- shence of systemic bias.
Abbreviations. EUR, European-Americans. AFR, African-Americans. AMR, Admixture
Americans Abbreviations. EUR, Euro
Americans.
Franciscus 529 Americans. EUR, EUR, EUR, Americans. African-Americans. African-Americans. African-Americans. Americans. Americans. Americans. African-Americans. African-Americans. African-Americans. African-Americans. Americans. American
- 530

It is made available under a [CC-BY 4.0 International license](http://creativecommons.org/licenses/by/4.0/) .

532 Figure 2. Rare variant association study in the European-American cohort and

533 replication of CR1 and APOL1 variants in the Admixed American cohort.

- 534(A) Shown graphically are the results of rare-variant association study involving the
- 535
- 536
- 537
- 538
- 539
- European-American subject cluster (gnomAD EUR AF < 0.01; missense and PTV (truncating variants). Statistical approaches included the following: the Simes me
for multiple hypothesis testing, the Fisher exact test for testin 540
- 541
- Truncating variants). Statistical approaches included the following: the Simes method
for multiple hypothesis testing, the Fisher exact test for testing two groups, C-alpha test
for comparing the variance of each group aga for multiple hypothesis testing, the Fisher exact test for testing two groups, C-alpha tes
for comparing the variance of each group against the expected, adaptive sum statistic
(ASUM) for testing variants; , weighted sum s for comparing the variance of each group against the expected, adaptive sum statistic
(ASUM) for testing variants; , weighted sum statistics (WSS) for testing variants and
kernel-based adaptive clustering (KBAC) for varian (ASUM) for testing variants; , weighted sum statistics (WSS) for testing variants and
kernel-based adaptive clustering (KBAC) for variant classification and association
testing. Of the top 10 (most significant) genes, fou kernel-based adaptive clustering (KBAC) for variant classification and association
testing. Of the top 10 (most significant) genes, four with known FSGS-associated vari
CR1, complement C3b/C4B receptor 1; *KANK1*, KN mot testing. Of the top 10 (most significant) genes, four with known FSGS-associated v:
CR1, complement C3b/C4B receptor 1; KANK1, KN motif and ankyrin repeat doma
COL4A4, collagen type 4, alpha 4 chain; IL-36G, interleukin 3
-

- 546
- 547
- 548
- CR1, complement C3b/C4B receptor 1; KANK1, KN motif and ankyrin repeat domains 1;
COL4A4, collagen type 4, alpha 4 chain; IL-36G, interleukin 36G.
Shown are associations of FSGS-related SNPs in CR1 and APOL1 in the GnomAD *GR1*, complement C3b/C4B receptor 1; *KANK1*, KN motif and ankyrin repeat domains 1;
 COL4A4, collagen type 4, alpha 4 chain; IL-36G, interleukin 36G.
 544
 545 (B) Shown are associations of FSGS-related SNPs in *CR COL4A4*, collagen type 4, alpha 4 chain; IL-36G, interleukin 36G.

544

545 (B) Shown are associations of FSGS-related SNPs in *CR1* and *APOL1* identabase population among European-Americans (top panels) a

547 (bottom p 545 (B) Shown are associations of FSGS-related SNPs in CR1 and APOL1 in the GnomAD aggregation
database population among European-Americans (top panels) and Admixed-Americans
(bottom panels). Left Panels. Two rare variant for the panels). Left Panels. Two rare variants in *CR1* have been associated with FSGS
Euro-Americans, and one of these variants in *CR1* has also been associated with FSGS in
Americans (both variants in *CR1* are common (bottom panels). Let Fancis. Two rare variants in CRI have been associated with FSGS in La
Euro-Americans, and one of these variants in CRI has also been associated with FSGS in La
Americans (both variants in CRI are commo Euro-Americans, and one of these variants in CR1 has also been associated with F505 in Eath-Americans (both variants in CR1 are common in Latin-Americans). **Right panels**. Two rare variants in *APOL1* are associated with F 549
- Americans (both variants in CR1 are common in Latin-Americans). **Right panels**. Two rare
variants in *APOL1* are associated with FSGS in European-Americans and in Latin-Americans. 550 variants in APOL1 are associated with FSGS in European-Americans and in Latin-Americans.

