Integrative Genomics Identifies Novel Associations with APOL1 Risk Genotypes in Black NEPTUNE Subjects

Matthew G. Sampson,* Catherine C. Robertson,** Sebastian Martini,† Laura H. Mariani,‡ Kevin V. Lemley,† Christopher E. Gillies,* Edgar A. Otto,* Jeffrey B. Kopp,§ Anne Randolph,† Virginia Vega-Warner,* Felix Eichinger,† Viji Nair,† Debbie S. Gipson,* Daniel C. Cattran,‖ Duncan B. Johnstone,¶ John F. O’Toole,** Serena M. Bagnasco,†† Peter X. Song,†† Laura Barisoni,§§ Jonathan P. Troost,* Matthias Kretzler,¶¶ John R. Sedor,**¶¶ and the Nephrotic Syndrome Study Network

Due to the number of contributing authors, the affiliations are listed at the end of this article.

ABSTRACT

APOL1 variants have been associated with renal phenotypes in blacks. To refine clinical outcomes and discover mechanisms of APOL1-associated kidney injury, we analyzed clinical and genomic datasets derived from 90 black subjects in the Nephrotic Syndrome Study Network (NEPTUNE), stratified by APOL1 risk genotype. Ninety subjects with proteinuria $\geqslant 0.5 \text{ g/d}$ were enrolled at first biopsy for primary nephrotic syndrome and followed. Clinical outcomes were determined, and renal histomorphometry and sequencing of Mendelian nephrotic syndrome genes were performed. APOL1 variants were genotyped, and glomerular and tubulointerstitial transcripts from protocol renal biopsy cores were analyzed for differential and correlative gene expression. Analyses were performed under the recessive model (high-risk genotype defined by two risk alleles). APOL1 high-risk genotype was significantly associated with a 17 ml/min per 1.73 m² lower eGFR and a 69% reduction in the probability of complete remission at any time, independent of histologic diagnosis. Neither APOL1 risk group was enriched for Mendelian mutations. On renal biopsy, high-risk genotype was associated with increased fractional interstitial area, interstitial fibrosis, and tubular atrophy. Risk genotype was not associated with intrarenal APOL1 mRNA expression levels. Differential expression analysis demonstrated an increased steady-state level of five genes associated with the high-risk genotype (CXCL9, CXCL11, and UBD in glomerulus; SNOR14B and MUC13 in tubulointerstitium). APOL1 tubulointerstitial coexpression analysis showed coexpression of APOL1 mRNA levels with a group of intrarenal transcripts that together were associated with increased interstitial fibrosis and tubular atrophy. These data indicate the high-risk APOL1 genotype confers renal risk across histopathologic diagnoses.

In 2010, two common exonic variants within apolipoprotein L1 (APOL1), termed G1 and G2 were associated with greatly increased risk of focal segmental glomerulosclerosis (FSGS) in blacks.¹ Subsequent epidemiologic studies have shown that harm associated with these APOL1 risk variants in blacks extends to renal phenotypes beyond FSGS.²⁻⁷ This includes nondiabetic CKD, HIV-associated nephropathy, progression to ESRD for patients with lupus nephritis or CKD, and development of poor renal outcomes over time in population controls.

The absence of the endogenous APOL1 gene in laboratory animal species has led to an increased importance of analyzing human data to understand the biologic mechanism underlying APOL1-related renal disease. Earlier work in humans has localized

Received November 24, 2014. Accepted May 31, 2015.
Published online ahead of print. Publication date available at www.jasn.org.

Correspondence: Dr. Matthew Sampson, Division of Pediatric Nephrology, University of Michigan School of Medicine, 8220D MSRB III, West Medical Center Drive, Ann Arbor, MI 48109. Email: mgsamps@med.umich.edu

Copyright © 2015 by the American Society of Nephrology
APOL1 to the podocyte in the glomerulus, the proximal tubule epithelium, and the arteriolar endothelium. APOL1 renal expression is reduced in subjects with CKD compared with healthy kidneys but is independent of the known risk variants, and ApoL1 plasma levels do not associate with kidney disease phenotypes. Transplantation studies indicate that the increased risk of allograft loss follows the kidney from the APOL1 risk allele-positive donor.

Studying APOL1-related kidney disease in the Nephrotic Syndrome Study Network (NEPTUNE) provides a unique opportunity to derive novel insights about the effect of these risk variants across proteinuric glomerular diseases. We genotyped 90 self-reported black subjects for the G1 and G2 risk alleles and classified these subjects with a low-risk (LR) or high-risk (HR) genotype. We report on the associations with clinical characteristics at enrollment and on clinical outcomes over time. Taking advantage of each subject’s paired molecular and histologic data, we also report on associations with genetic, gene expression, and histologic parameters.

RESULTS

APOL1 Allele and Genotype Frequency in NEPTUNE Subjects

The minor allele frequencies for G1 and G2 were 0.32 and 0.23, respectively (Supplemental Table 1). The two risk allele (HR) genotype was considered any combination of G1 and G2 and was found in 43% of subjects. In a population-based study of blacks, the two risk allele genotype frequency was 13%. The APOL1 risk allele frequency in our subjects was significantly elevated in the FSGS and other glomerulopathy cohorts, but not in those with minimal change disease (MCD) or membranous nephropathy (MN) (Table 1).

Genetic Model for Association Studies

For all association studies, we used the recessive model of risk inheritance as the primary model: 0 or 1 risk alleles (LR) versus HR genotype.

Table 1. Frequency of HR genotype in NEPTUNE versus Dallas Heart Study

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Frequency of HR Genotype</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dallas Heart Study</td>
<td>0.13 (236/1776)</td>
<td>ref</td>
</tr>
<tr>
<td>NEPTUNE all</td>
<td>0.43 (39/90)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>NEPTUNE FSGS</td>
<td>0.60 (25/42)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>NEPTUNE MCD</td>
<td>0.25 (5/20)</td>
<td>0.17</td>
</tr>
<tr>
<td>NEPTUNE MN</td>
<td>0.09 (1/11)</td>
<td>1.00</td>
</tr>
<tr>
<td>NEPTUNE other</td>
<td>0.47 (8/17)</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Proportion of subjects with two risk alleles (HR genotype) in NEPTUNE by histologic diagnosis as compared with a population cohort of blacks (Dallas Heart Study). The P values were determined using Fisher’s exact test. Other, other glomerulopathies; ref, reference.

Monogenic Forms of Steroid-Resistant Nephrotic Syndrome

Causative mutations in known monogenic steroid-resistant nephrotic syndrome genes were not found in any individuals (see Supplemental Table 2 for genes sequenced). The frequency of rare, heterozygous, protein-altering variants in known recessive steroid-resistant nephrotic syndrome genes did not differ by risk genotype.

Subject Genetic Ancestry

Genotype-based ancestry determination using principal component analysis methods showed that 80 of the 82 self-reported black subjects with genome-wide data clustered with the African continental ancestry group of 1000 Genomes Project Subjects (Supplemental Figure 1). The remaining two subjects deviated further from this cluster, indicative of greater admixture.

Baseline Characteristics

Table 2 summarizes selected characteristics of the 90 black subjects stratified by LR versus HR. Treating clinicians of these subjects were not aware of the APOL1 status. There was no significant difference between LR and HR subjects in demographic characteristics, duration of disease before onset, duration of follow-up, administration of immunosuppressive medication, or renin-angiotensin blockade.

There was a significant difference in the distribution of the histopathologic diagnoses between the HR and LR groups. The largest differences were a higher prevalence of FSGS and a lower prevalence of MN in the HR group. The median eGFR at baseline was significantly lower in the HR group, at 67 versus 84 ml/min per 1.73 m² (P=0.02).

Of the 90% of subjects providing birth history information, 17% (14/81) were preterm (Supplemental Table 3). Nine of 14 preterm subjects were in the HR group. FSGS was the histologic diagnosis in eight of the nine HR preterm subjects. There was an increased, but not statistically significant, odds of prematurity associated with the HR genotype (3 [95% confidence intervals, 0.8–12.7], P=0.08). There was no significant difference in birthweight between the HR and LR groups, at 2856 versus 3176 g, respectively (P=0.10).

Analysis of Primary Clinical Outcomes

Given its deleterious effects across renal phenotypes, our primary hypothesis was that the HR APOL1 genotype is associated with worse clinical outcomes in all NEPTUNE diagnoses. We therefore studied the full cohort rather than stratifying by histopathologic diagnosis (histopath). We did include histopath as a covariate in our multivariate models. Of note, an interaction term between risk-genotype status and histopath was evaluated in each model. It was not significant in any of the models tested. However, our power calculations suggest we are underpowered to detect interactions (Supplemental Appendix 1).

eGFR

Fitting a multivariable linear regression using generalized estimating equations (GEEs), the HR genotype was significantly
associated with a 17 ml/min per 1.73 m² lower eGFR (<0.01) (Table 3). For those subjects whose first and last serum creatinine measurements were taken at least 8 months apart, we also analyzed the rate of change in eGFR over time. We computed the slope of the least-squares regression line of eGFR against time for each subject whose first and last serum creatinine levels were obtained at least 8 months apart (n=73). Across these subjects, the eGFR decreased modestly over time (approximately 8 ml/min per y, <0.01). However, this was not associated with risk genotype in unadjusted or adjusted models (Supplemental Appendix 2, Supplemental Table 4).

Urine Protein to Creatinine Ratio

A multivariable linear regression using GEEs showed a 52% increase in the urine protein to creatinine ratio (UPC) in HR subjects (P=0.06) (Table 4).

Complete Remission

Of the subjects, 49% of LR versus 23% of HR subjects achieved complete remission (CR) of proteinuria (P=0.02)

Histomorphometric Differences between LR and HR Status

Average glomerular volume, fractional interstitial area, and cortical glomerular density were measured in the 22 subjects with available slides (LR: n=11, HR: n=12). Additionally, the percentages of interstitial fibrosis (IF) and tubular atrophy (TA) were measured in 63 subjects. Clinical characteristics of these subgroups are provided in Supplemental Table 7. Although average glomerular volume and cortical glomerular density did not show significant association with genotype, those with HR genotypes had a significantly increased percentage of fractional interstitial area (25% versus 13%, P<0.01) (Figure 2, Supplemental Table 8). Percentages of IF and TA were elevated in HR subjects (Supplemental Table 9). Average IF was 9.4% in LR subjects versus 22.4% in the HR group (P<0.01). Similarly, average TA was 10.5% versus 23.6% in the LR and HR groups, respectively (P=0.01).

Glomerular and Tubulointerstitial Gene Expression

Intrarenal gene expression studies were performed using manually dissected glomeruli (GLOM) and tubulointerstitium (TI) from 44 and 55 subjects, respectively. These sample sizes reflect the biopsy samples that were available for expression analysis and which passed subsequent quality control measures at the time of study. Availability of samples was not biased toward any demographic or clinical metrics (Supplemental Table 7). As seen in the parent cohort, those in the HR subgroup had point estimates that were lower for eGFR and higher for percentage of patients that were FSGS. For both differential expression and correlation analyses, we first applied a variance filter to our gene set, reducing the pool of genes for testing to 3796 in GLOM and 3761 in TI (see the methods section for details).

Table 2. Characteristics of 90 black NEPTUNE subjects

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>LR (n=51)</th>
<th>HR (n=39)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y)</td>
<td>17.0 (12.0–46.5)</td>
<td>17.0 (14.5–32.5)</td>
<td>0.80</td>
</tr>
<tr>
<td>Pediatric</td>
<td>27 (53)</td>
<td>21 (54)</td>
<td>1.00</td>
</tr>
<tr>
<td>Age of onset (y)</td>
<td>15.5 (10.0–40.8)</td>
<td>17.0 (14.0–28.2)</td>
<td>0.60</td>
</tr>
<tr>
<td>Duration disease (y)</td>
<td>1.0 (0.0–3.0)</td>
<td>0.0 (0.0–1.8)</td>
<td>0.40</td>
</tr>
<tr>
<td>Male</td>
<td>32 (63)</td>
<td>25 (64)</td>
<td>1.00</td>
</tr>
<tr>
<td>Histopath</td>
<td></td>
<td></td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>FSGS</td>
<td>17 (33)</td>
<td>25 (64)</td>
<td></td>
</tr>
<tr>
<td>MCD</td>
<td>15 (29)</td>
<td>5 (13)</td>
<td></td>
</tr>
<tr>
<td>MN</td>
<td>10 (20)</td>
<td>1 (3)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>9 (18)</td>
<td>8 (21)</td>
<td></td>
</tr>
</tbody>
</table>

Clinical characteristics at baseline

| RAAS blockade                  | 14 (27)  | 11 (28)  | 1.00    |
| eGFR                           | 83.7 (65.8–105.6) | 67.2 (42.9–84.5) | 0.02    |
| UPC                            | 2.9 (1.9–7.2) | 3.1 (1.3–6.9) | 0.94    |

Data presented as median (interquartile range) for quantitative variables and as count (%) for categorical characteristics. The P values were determined using Wilcoxon rank-sum test or Fisher’s exact test. Age, patient age at enrollment; Other, other glomerulopathies.

Table 3. Multivariate GEE analysis of eGFR and APOL1 risk status adjusting for histopath, age, baseline UPC, and RAAS blockade

<table>
<thead>
<tr>
<th>Covariate</th>
<th>β</th>
<th>95% CI</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR</td>
<td>−17.2</td>
<td>−29.0 to 5.4</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Histopath</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSGS</td>
<td>ref</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>MCD</td>
<td>22.0</td>
<td>6.7 to 37.2</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>MN</td>
<td>8.5</td>
<td>−8.9 to 25.9</td>
<td>0.34</td>
</tr>
<tr>
<td>Other</td>
<td>−1.4</td>
<td>−15.8 to 13.1</td>
<td>0.85</td>
</tr>
<tr>
<td>Age (y)</td>
<td>−1.1</td>
<td>−1.4 to −0.7</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>UPC baseline</td>
<td>−0.8</td>
<td>−1.6 to 0.0</td>
<td>0.04</td>
</tr>
<tr>
<td>RAAS</td>
<td>2.6</td>
<td>−3.7 to 8.9</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Parameter estimates and P values determined using GEEs. Other, other glomerulopathies; Age, patient age at enrollment; ref, reference; 95% CI, 95% confidence interval.

APOL1 Differential Expression

APOL1 HR status was not significantly associated with APOL1 expression in the GLOM (fold change, 1.0; P=0.65) or TI (fold change, 1.1; P=0.40) (Figure 3).
Table 4. Multivariate GEE analysis of repeated log2 UPC and APOL1 risk status adjusting for histopath, baseline eGFR, and RAAS blockade

<table>
<thead>
<tr>
<th>Covariate</th>
<th>β</th>
<th>2βa</th>
<th>95% CI</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR</td>
<td>0.6</td>
<td>1.5</td>
<td>-0.0 to 1.2</td>
<td>0.06</td>
</tr>
<tr>
<td>Histopath</td>
<td>ref</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>FSGS</td>
<td>-0.7</td>
<td>0.6</td>
<td>-1.8 to 0.4</td>
<td>0.22</td>
</tr>
<tr>
<td>MCD</td>
<td>1.3</td>
<td>2.5</td>
<td>0.7 to 1.9</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>MN</td>
<td>-0.1</td>
<td>0.9</td>
<td>-0.8 to 0.6</td>
<td>0.72</td>
</tr>
<tr>
<td>Other</td>
<td>-0.6</td>
<td>0.7</td>
<td>-1.2 to -0.1</td>
<td>0.03</td>
</tr>
<tr>
<td>RAAS baseline</td>
<td>-0.0</td>
<td>1.0</td>
<td>-0.0 to -0.0</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Parameter estimates and P values determined using GEEs. Other, other glomerulopathies; ref, reference; 95% CI, 95% confidence interval.

a2β is the factor by which UPC increases in HR subjects compared with LR subjects. For example, if 2β is 1.5, the average UPC in HR subjects is 1.5 times that in LR subjects.

The increased IF and TA in the HR subjects raised the question of whether the association between TI APOL1 expression and risk genotype is confounded by differential degree of TI injury. IF and TA were available for 42 genotyped TI samples. There was no significant association between TI APOL1 expression and IF or TA (Supplemental Appendix 3).

There was no analogous measure of glomerular damage, however, as detailed in the methods section, only GLOM with open Bowman’s space are isolated across samples. Even when only studying subjects with FSGS, there were no significant differences in GLOM APOL1 expression between the HR and LR groups (Figure 3).

Figure 1. Unadjusted Kaplan–Meier plot of proportion of CR stratified by APOL1 risk status. Plot is truncated at day 1000. P value determined with an unadjusted Cox proportional hazards model of CR and APOL1 risk status.

Genome-Wide Differential Expression

Three GLOM and two TI genes had significantly higher expression levels in HR subjects (FDR<0.05) (Tables 5 and 6). The most significantly differentially expressed gene was CXCL9 in GLOM, whose average expression in HR subjects was 4.1 times that of LR subjects. Expression was also higher in GLOM for another member of the CXCL family, CXCL11, and the proteosomal degradation pathway gene, ubiquitin D (UBD) also known as FAT10. Small nucleolar RNA 14B and mucin 13 (MUC13) transcript levels were significantly higher in the HR TI compartment. Quantile-quantile plots, volcano plots, and boxplots of expression by risk genotype for significant genes are provided in Supplemental Note 5.

APOL1 Coregulation Analysis

Condordant changes in mRNA transcript levels can identify gene sets that underlay a unified regulatory mechanism and function. In fact, mRNA coexpression patterns can result from common upstream regulatory events that determine the transcriptional response in the functional context of disease.13–15 Although our data indicate that APOL1 risk genotypes do not influence the magnitude of its expression, we hypothesized that these exonic variants alter the function of the APOL1 protein and result in changes in the expression of other genes.

Identifying sets of genes whose expression is differentially correlated with HR versus LR APOL1 could help elucidate the biologic mechanism by which APOL1 risk variants contribute to disease. To model coexpression of genes with the HR and LR APOL1 transcripts, we used a Weighted Gene Correlation Network Analysis (WGCNA).16 In the WGCNA, a gene expression network is constructed using coexpression relationships in a weighted manner. Distinct coregulated gene sets (modules) are identified within this network. The genes contained in these modules can then be explored for biologic concepts. We used the WGCNA in TI to identify the coexpression network of the HR APOL1 transcript and compared it with that of the LR transcript. For this analysis, we adjusted for age, sex, and histologic diagnosis.

In the analysis using TI data from only HR subjects, the APOL1 transcript was part of a coexpression module containing 579 genes. This module’s expression was positively correlated with tissue injury (IF: r=0.43, P=0.04; TA: r=0.42, P=0.05). When analyzing LR subjects, TI APOL1 was part of a 681-gene module. Although the point estimate showed a positive correlation of lower magnitude with IF and TA, neither was statistically significant (IF: r=0.33, P=0.06; TA: r=0.3, P=0.09).

We next evaluated molecular pathways, concepts, and upstream regulators associated with the gene modules containing TI APOL1 from the HR and LR group analyses. Both HR and LR modules had overlapping canonical pathways and predicted upstream regulators and many innate immunity, adaptive immunity, and immune costimulation pathways. However, the two modules also had a substantial set of distinct pathways and
regulators (Supplemental Figure 2, Supplemental Table 10). For instance, the HR module was specifically enriched for the pathways of role of Jak family kinases in IL-6-type cytokine signaling, ILK signaling, and JAK/Stat signaling.

**DISCUSSION**

Through its longitudinal design and collection of detailed and diverse data and biosamples, the NEPTUNE provided a unique laboratory to discover molecular mechanisms underlying the pathogenicity of APOL1 in human subjects with proteinuric glomerular diseases. In addition, we used this unique cohort of deeply phenotyped individuals to more fully plumb the depths of the clinical correlates of this risk genotype.

This study was designed to test the hypothesis that, independent of classic histopath (e.g., MCD, FSGS, MN), the HR genotype resulted in worse phenotypic outcomes and a distinct intrarenal gene expression profile. In fact, we demonstrated that, in our combined cohort of 90 subjects, the HR genotype was associated with a lower eGFR and a lower likelihood of attaining CR. In multivariable analyses, these differences were not attributable to differences in age, sex, duration of disease, or duration of follow-up time.

Although not the primary question of this study, a distinct and interesting question is whether the HR genotype had a differential effect, depending on specific histopathologic diagnoses. Although underpowered, we performed this stratified analysis for eGFR, UPC, and CR. In regard to eGFR and CR, these exploratory analyses showed the same direction of effect associated with the HR genotype in FSGS, MCD, and other strata. For UPC, subjects with the HR genotype and MCD had a lower UPC than the other groups. In other cohorts, or as NEPTUNE expands its enrollment, future studies with adequate power can be specifically designed to study the HR genotype in distinct glomerular conditions.

The results of the primary analyses, and the suggestive support of stratified analyses in terms of direction of effect, support the concept that the HR genotype confers risk across histopathologic diagnoses. Even if HR genotype frequency is not enriched in MCD, MN, or other glomerulopathies, those who have it may represent a more vulnerable population. This seems to be an important area of further research because the field seeks to determine those populations and renal phenotypes in which there is clinical utility in ascertaining the APOL1 genotype.

Other studies have found that HR subjects have a more rapid decline in eGFR compared with those without it. In regression analysis of the rate of change in eGFR adjusted for its baseline levels, we did not observe this pattern. It may be that insufficient time had passed at time of analysis to discern differences in eGFR trajectory between groups. To sort this out, it will be necessary to continue to follow this over time in the NEPTUNE and in other clinical studies with similar intensity of follow-up.

HR subjects had significantly more IF and TA on biopsy at enrollment. Worse histologic metrics of renal tubulointerstitial disease with the HR genotype were also recently reported in a group of blacks with non-nephrotic CKD. Increased TI damage may, in part, explain the eGFR differences observed. The lack of difference in prior duration of disease implies that, from the outset, an HR kidney’s response to challenge or injury is accompanied by an increased inflammatory response. There were no quantitative differences in glomerular morphometry at the time of initial biopsy. It will be important to discern whether glomerular histomorphometric differences are present at baseline with effect sizes too small to detect at this sample size or if they only become apparent later in the disease course.

The HR genotype did not alter APOL1 steady-state mRNA levels in either compartment. A previous report showed that the HR genotype is not associated with circulating APOL1
Together, this provides additional support for the concept that the G1 and G2 genotypes of \textit{APOL1} result in damage by a mechanism other than altering their overall expression. Potential mechanisms could be through altering the protein's overall conformation or the structure of critical epitopes.

\text{CXCL9}, a leukocyte secreted chemokine induced by IFN-\(\gamma\), showed the highest risk genotype-specific expression difference in GLOM, with a fold change of 4.1 in HR versus LR subjects. Along with \text{CXCL11} (also with higher expression in the HR subjects), this chemokine mediates leukocyte migration to areas of tissue damage and has been described in intrinsic renal cell lineages.\textsuperscript{18} \textit{UBD} was significantly increased in GLOM of HR subjects. Upregulated by INF-\(\gamma\) and TNF-\(\alpha\), \textit{UBD} targets proteins for proteasomal degradation by a ubiquitin-independent mechanism\textsuperscript{19,20} and can induce apoptosis.\textsuperscript{21} In addition to its role in innate immunity, the function of \textit{APOL1} as a proapoptotic and autophagic cell death gene has been described.\textsuperscript{22} It is possible that increased \textit{UBD}-mediated proteasomal degradation is one mechanism that could link an aggravated immune response in HR \textit{APOL1} subjects with downstream renal damage. \textit{MUC13}, a transmembrane mucin gene, was more highly expressed in TI. \textit{MUC13} has a role in protection of epithelial cells from pathogen infection and also functions in cell signaling.\textsuperscript{23}

Correlated expression levels among genes in a disease- and compartment-specific manner have been used to identify functional relationships, including physical interactions and regulatory mechanisms.\textsuperscript{24} For the HR and LR genotypes separately and independent of histologic diagnosis, we found a single coexpression module of genes that included the expression of \textit{APOL1} in TI. The HR module was significantly correlated with increased tissue injury. This gene set's functions were then analyzed \textit{in silico}, revealing enrichment of specific immune canonical pathways. We also identified potential regulators of these HR correlated genes. These findings fit in well with emerging concepts of drivers of \textit{APOL1}-mediated toxicity across renal phenotypes,\textsuperscript{25,26} while also providing new molecular targets for further investigation.

Through differential and intercorrelation analyses, we found sets of intrarenal transcripts associated with the HR genotype. Our studies do have several limitations. We showed that some of these transcripts, such as \textit{MUC13}, are associated with the HR genotype, independent of tissue injury. However, not all coexpressed or differentially expressed genes will be mediators of the damage caused by the HR genotype. Alternatively, the increased tissue injury associated with the HR genotype may itself result in differential regulation or coexpression of sets of transcripts. Whether identified as potential mediators of the association between genotype and tissue damage or a response to the HR genotype-mediated tissue damage, each of these transcripts can serve to illuminate important molecular mechanisms involved in the renal damage associated with \textit{APOL1}. In addition, the HR gene models identified by the WGCNA will require additional empirical confirmation. For example, protein-based assays and functional studies in model systems will be essential to validate the associations that we have observed. This vital step will also allow to more confidently delineate those causal pathways in these relationships.

### Table 5. Differentially expressed genes in glomerulus

<table>
<thead>
<tr>
<th>Gene</th>
<th>Fold Change</th>
<th>SAM Score</th>
<th>q Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CXCL9</td>
<td>4.1</td>
<td>2.9</td>
<td>0</td>
</tr>
<tr>
<td>CXCL11</td>
<td>2.7</td>
<td>2.4</td>
<td>0</td>
</tr>
<tr>
<td>UBD</td>
<td>3.0</td>
<td>2.4</td>
<td>0</td>
</tr>
</tbody>
</table>

SAM Score, \(T\) statistic value computed using Significance Analysis of Microarrays \(q\) value, lowest false discovery rate at which the gene is called significant.
Our study also did not have a measurement of tissue injury from each patient in whom expression analyses were performed, which prevented us from adjusting for tissue injury in each subject when relating HR genotype to gene expression. As the NEPTUNE cohort matures, each subject will have morphologic descriptors of tissue injury measured, which will allow us to account for this specific covariate in our models. Finally, we were underpowered to stratify by standard pathologic diagnosis. However, as the cohort expands, we will be empowered to perform expression analyses stratified by conventional diagnoses.

Using integrative genomics, we have shown that individuals with the HR APOL1 genotype enrolled in the NEPTUNE study present, from a clinical and histologic perspective, with more advanced disease and have less remission of proteinuria over time. This appears to be independent of histopath. If independently confirmed, these observations may imply that defining a patient’s APOL1 genotype could allow better stratification of subjects for trials which could eventually result in more precise and improved clinical management. Finally, transcriptomic data suggest that intrarenal APOL1 functions in the context of pathways of innate immunity and inflammation. It appears that genotype-associated expression differences reflect an aggravated inflammatory response to a challenge rather than activation of a de novo APOL1-specific program. Combining these emerging human tissue–based hypotheses with further longitudinal analyses and investigations of candidate molecules in experimental model systems should aid in defining the causal mechanism of APOL1.

### CONCISE METHODS

#### Study Participants

There were 92 subjects of self-reported black ancestry from the NEPTUNE. Two subjects did not give consent for genetic analyses and were not included in the genetic and transcriptomic analyses. The design of the NEPTUNE has been previously described. Subjects were recruited at the time of their first clinically indicated biopsy for suspicion of primary proteinuric glomerular disease. Study pathologists classified subjects into MCD, FSGS, MN, and other glomerulopathies histopathologic groups as previously defined.

#### APOL1 Genotyping

The APOL1 SNPs corresponding to G1 (rs73885319 and rs60910145) and the G2 indel (rs717185313) were genotyped using Sanger sequencing of the last 250 bases of exon 7. Individuals were classified as having 0, 1, or 2 risk genotypes. The rare G+ allele, with rs73885319 being variant and rs60910145 being reference, was classified as a risk allele. There was no distinction made between G1 or G2 in subjects with 1 or 2 risk genotypes.

#### Mendelian FSGS Gene Sequencing

Using established protocols and gene targets, targeted multiplex PCR paired with next-generation sequencing was used to determine if blacks had rare, deleterious alleles in 20 genes known to cause Mendelian forms of FSGS. A deleterious allele was defined as meeting the following three criteria: (1) allele frequency <0.1% for dominant genes or <0.5% for recessive genes, (2) present in <5% of our cohort, and (3) a PolyPhen2 score of damaging. The R229Q polymorphism of podocin was also considered in compound heterozygosity with rare podocin mutations.

#### Clinical Characteristics at Presentation

Data are summarized as mean±SD. Fisher’s exact test was used to compare count characteristics of individuals with the LR versus HR genotypes. For numerical variables, Wilcoxon rank-sum tests were used to test for differences in distributions between the two groups. Fisher’s exact test was used to assess the association between risk genotype and preterm status.

#### Longitudinal Analysis of eGFR and UPC

eGFR was calculated using the Schwartz formula for the pediatric population and the Chronic Kidney Disease Epidemiology Collaboration equation for adults. UPC was calculated from a random collection on a milligram/milliliter ratio.

Two generalized estimating equation linear regressions were fit to assess the association of eGFR and UPC with risk genotype over the duration of subjects’ follow-up. Covariates to include in the models were chosen on the basis of a priori determination of clinically meaningful factors. The GEE analysis of eGFR was adjusted for histopath, age, baseline UPC, and use of renin-angiotensin aldosterone system (RAAS) blocker. The GEE analysis of UPC was adjusted for histopath, RAAS blocker, and baseline eGFR. In addition to fitting for marginal effect models, we also explicitly tested for interaction between histopath and risk genotype in both.

An analysis of the effect of risk genotypes on the rate of decline in eGFR over time was also performed. First, the slope of the least-squares regression line of eGFR versus time was calculated in all subjects whose eGFR over time was also performed. Then, two linear regressions of the eGFR slope by risk genotype were fit: one adjusting for baseline eGFR, age, baseline UPC, RAAS blocker, and histopath and another adjusting for baseline eGFR only.

#### Time to Event Analysis of CR

A Cox proportional hazards model was used to determine association between time to first CR of proteinuria and APOL1 risk genotype, adjusting for cohort and baseline age, UPC, and eGFR. Model selection was performed using the same approach as previously described.

#### Power and Stratified Analyses

Analogous analyses of eGFR, UPC, and CR were also performed stratified by histologic cohort. Because of the small sample sizes,

<table>
<thead>
<tr>
<th>Table 6. Differentially expressed genes in T1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>MUG13</td>
</tr>
<tr>
<td>SNORA14B</td>
</tr>
</tbody>
</table>

SNORA14B, small nucleolar RNA 14B; SAM Score, T statistic value computed using Significance Analysis of Microarrays; q value, lowest false discovery rate at which the gene is called significant.
stratified longitudinal models were considered exploratory in nature. Simplified a priori power calculations for both combined and stratified analyses were performed. Details of the methods used to estimate power are described in Supplemental Appendix 1, and the results are presented in Supplemental Appendix 2.

**Histomorphometric Measurement**

During annotation of biopsies, individual GLOM were identified throughout all sections in which they appeared. Measurements were made using the SlidePath platform (Leica Microsystems, Dublin) by a single individual (K.L.). Average glomerular tuft area was determined in a single section, usually one near the middle of the stack of slides, usually with periodic acid–Schiff staining. The number of patent tufts in the section was counted and was divided by the cortical area of the section. This was done on one section each stained with periodic acid–Schiff and with Masson’s trichrome stain. Less than 125 points per stained section were rarely counted, if the amount of tissue was insufficient for five distinct cortical locations to be chosen.

**Quantitation of IF and TA**

Digital images of the biopsies were examined for scoring by pathologists, and for each biopsy whole-slice images of glass slides were scanned using Aperio technology at 40× for light microscopy. The degree of cortical IF and TA was scored as the percentage of cortex by each reader, and the average of the scores from at least two readers was used.

**Gene Expression**

**RNA Preparation**

Renal biopsy–derived gene expression data were collected and analyzed from subjects who had APOL1 genotype data available. There were 44 subjects with glomerular (LR: n=29; HR: n=15), and 55 subjects with tubulointerstitial transcriptomic data (LR: n=32; HR: n=23). The renal biopsy sample was manually dissected and isolated into tubulointerstitial and glomerular compartments per established protocol.

In terms of glomerular injury, manual glomerular microdissection captures GLOM with an open Bowman’s space. Therefore, across samples, our GLOM expression studies are enriched for the transcriptomes of functioning GLOM, rather than for those that are globally sclerosed. RNA was purified according to the manufacturer’s instructions.

**Microarray Expression**

Renal biopsy specimens were processed and analyzed using Affymetrix 2.1 ST chip as previously described. Gene expression was normalized and quantitated on a probeset level as previously described.

The 44 GLOM and 55 TI expression datasets generated are being uploaded to the Gene Expression Omnibus. These data are stratified by APOL1 risk genotype status and are available at the time of publication. All analyses were performed using R software.

**APOL1 Expression by Risk Genotype**

Using a t test, we compared microarray-based gene-level expression of APOL1 between individuals with the LR versus HR genotypes. Expression levels were log2 transformed.

**Genome-wide Expression by Risk Genotype**

To identify differentially expressed genes in HR versus LR cohorts, we first calculated the variance of expression levels for all probesets on the array. This was done in a GLOM- and TI-specific manner. We subsequently restricted the set of transcripts analyzed to those corresponding to the 15% of probesets with the largest amount of variation in expression levels across samples. This resulted in 4012 transcripts. Differential expression analysis was then performed on these transcripts using Significance Analysis of Microarrays methods to properly account for multiple testing and correlation among genes. In this approach, permutations of the observed data are used to determine the expected distribution of test statistics under the assumption of no association between expression and APOL1 genotype. Then a significance threshold is empirically determined such that FDR is less than a desired value. In this framework, FDR is defined as the median number of genes called significant across the permutations divided by the number of genes called significant in the observed data. Extensive documentation of the statistical methods and computational tools used are provided in the Significance Analysis of Microarrays user guide (http://statweb.stanford.edu/~tibs/SAM/sam.pdf) in addition to the cited article. The analysis underwent 1000 permutations, and an FDR threshold of 0.05 was used to determine statistical significance.

Power calculations for this analysis were performed. For genes with a pooled SD≤0.9, we had an estimated 80% power to detect fold changes of 3.2 and 2.7 in glomerular and tubular expression, respectively (Supplemental Appendix 4).

**APOL1-correlated Expression**

Identifying genes coexpressed with APOL1 and performing functional annotation can aid in illuminating the mechanism of this molecule in the LR and HR genotypes. After adjusting gene expression values for age, sex, and histopath, and again restricting our analyses to probesets with the most variation across samples, we used the published method weighted gene correlation analysis and its corresponding R package WGCNA.

For the HR transcript analysis, the following parameters were used: (1) sample cut height=83, (2) soft thresholding parameter=12, and (3) merge module cut height=0. The LR transcript analysis parameters are as follows: (1) sample cut height=95, (2) soft thresholding parameter=7, and (3) merge module cut height=0. On visual inspection, none of the samples in the HR group for GLOM or TI appeared to be a substantial outlier. Therefore, the sample cut height was set to be just larger than the maximum height observed in the dendogram. Meanwhile, both dendograms for GLOM and TI expression in the LR group demonstrated one
subject that was significantly separated from other samples. The selected cut height of 95 resulted in these samples being excluded from the analysis. The soft thresholding parameter for each analysis was chosen to achieve a scale free topology model fit ($R^2$) $\geq$0.9 and such that a reasonable number of distinct gene modules were identified from the gene network (between 8 and 18). A merge module cut height of 0 was chosen to preserve the modules identified using the Dynamic Tree Cut (see WGCNA documentation for more details http://labs.genetics.ucla.edu/horvath/CoexpressionNetwork/Rpackages/WGCNA).

Using this method, a gene expression network was constructed in a manner designed to optimize biologic relevance, and then distinct gene sets were identified within the network. Association between the resulting gene sets and other traits were then tested using the eigengene of each gene set.

**Functional Analysis**

Gene sets emerging from correlation analysis using WGCNA were uploaded to Ingenuity Pathway Analysis (QIAGEN, Redwood City, CA) for functional annotation.

**Statistical Analyses**

All analyses were performed using R software. Additionally, two-sided tests of hypotheses were used, and $P<0.05$ was the criterion for statistical significance unless otherwise specified.

**ACKNOWLEDGMENTS**

MS is supported by 1K08-DK100662-01 and MK is supported by U54-DK083912 Nephrotic Syndrome Study Network Consortium and P30-DK081943 George M. O’Brien Kidney Research Core Center at the University of Michigan. The Nephrotic Syndrome Study Network Consortium (NEPTUNE), U54-DK-083912, is a part of the National Center for Advancing Translational Sciences (NCATS) Rare Disease Clinical Research Network (RDCRN), supported through a collaboration between the Office of Rare Diseases Research (ORDR), NCATS, and the National Institute of Diabetes, Digestive, and Kidney Diseases. RDCRN is an initiative of ORDR, NCATS, and the National Institute of Diabetes, Digestive, and Kidney Diseases. Parts of this study were presented at the 2014 American Society of Nephrology Annual Meeting, November 13–16, in Philadelphia, PA.

**DISCLOSURES**

None

**REFERENCES**


This article contains supplemental material online at http://jasn.asnjournals.org/lookup/suppl/doi:10.1681/ASN.2014111131/-/DCSupplemental.

AFFILIATIONS

*Division of Nephrology, Department of Pediatrics and Communicable Diseases, †Division of Nephrology, Departments of Internal Medicine and Computational Medicine and Bioinformatics, and ‡Division of Computational Medicine and Bioinformatics, University of Michigan School of Medicine, Ann Arbor, Michigan; ‡‡Division of Nephrology, Department of Pediatrics, Children’s Hospital of Los Angeles, University of Southern California School of Medicine, Los Angeles, California; §Division of Nephrology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland; ‡‡‡Division of Nephrology, University Health Network, University of Toronto, Toronto, Ontario, Canada; ††Division of Nephrology, Department of Internal Medicine, Temple University School of Medicine, Philadelphia, Pennsylvania; **Division of Nephrology, Department of Internal Medicine and †††Division of Physiology and Biophysics, Case Western Reserve University and Rammelkamp Center for Education and Research, MetroHealth System, Cleveland, Ohio; ††††Department of Pathology, Johns Hopkins School of Medicine, Baltimore, Maryland; ‡‡‡‡Division of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan; and ‡‡‡‡‡Division of Pathology, University of Miami, Miller School of Medicine, Miami, Florida