Whole-Genome Linkage and Association Scan in Primary, Nonsyndromic Vesicoureteric Reflux

Heather J. Cordell,* Rebecca Darlay,* Pimphen Charoen,†‡ Aisling Stewart,* Ambrose M. Gullett,§ Heather J. Lambert,† Sue Malcolm,§ Sally A. Feather,¶ Timothy H.J. Goodship,* Adrian S. Woolf,§ Rajko B. Kenda,** and Judith A. Goodship,* for the UK VUR Study Group

*Institute of Human Genetics, Newcastle University, Newcastle upon Tyne, United Kingdom; †Department of Medical Genetics, University of Cambridge, Cambridge, United Kingdom; ‡Department of Tropical Hygiene, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand; §UCL Institute of Child Health, University College, London, United Kingdom; ‖Royal Victoria Infirmary, Newcastle upon Tyne, United Kingdom; ¶St. James’ University Hospital, Leeds, United Kingdom; and **Department of Pediatric Nephrology, University Medical Centre Ljubljana, Ljubljana, Slovenia

ABSTRACT
Primary vesicoureteric reflux accounts for approximately 10% of kidney failure requiring dialysis or transplantation, and sibling studies suggest a large genetic component. Here, we report a whole-genome linkage and association scan in primary, nonsyndromic vesicoureteric reflux and reflux nephropathy. We used linkage and family-based association approaches to analyze 320 white families (661 affected individuals, generally from families with two affected siblings) from two populations (United Kingdom and Slovenian). We found modest evidence of linkage but no clear overlap with previous studies. We tested for but did not detect association with six candidate genes (AGTR2, HNF1B, PAX2, RET, ROBO2, and UPK3A). Family-based analysis detected associations with one single-nucleotide polymorphism (SNP) in the UK families, with three SNPs in the Slovenian families, and with three SNPs in the combined families. A case-control analysis detected associations with three additional SNPs. The results of this study, which is the largest to date investigating the genetics of reflux, suggest that major loci may not exist for this common renal tract malformation within European populations.


Vesicoureteric reflux (VUR) is abnormal movement of urine from the bladder retrogradely through the vesicoureteric junctions into the upper urinary tract. This is a study of primary VUR, i.e., VUR that is not secondary to bladder outflow obstruction caused by neurogenic damage or urethral valves or part of a multiorgan syndrome. VUR is usually a benign condition but can be associated with transient kidney damage, acute inflammation from ascending pyelonephritis, or permanent damage as a consequence of scarring after infection and/or congenital kidney defects histologically comprising renal hypoplasia (too few nephrons) and/or renal dysplasia (incomplete differentiation).1–3 These renal defects are grouped under the term reflux nephropathy (RN). In the United Kingdom, RN accounts for 12% of the approximately 40,000 adults and 7% of the 768 children who require renal transplantation and/or life-long dialysis.4

Received June 17, 2009. Accepted September 24, 2009.
Published online ahead of print. Publication date available at www.jasn.org.

Correspondence: Dr. Judith A. Goodship, Institute of Human Genetics, Newcastle University, International Centre for Life, Central Parkway, Newcastle upon Tyne, NE1 3BZ, UK. Phone: +44-0-191-241-8747; Fax: +44-0-191-241-8666; E-mail: j.a.goodship@ncl.ac.uk

Copyright © 2010 by the American Society of Nephrology
Traditionally, the diagnosis of VUR has been based on cystography with radiodense or radioisotopic materials to visualize retrograde passage of urine. Williams et al. reviewed 15 cystography studies in well children: The largest study reported no VUR in 722 children, whereas some of the smaller studies reported much higher percentages of affected individuals. The true prevalence of (primary) VUR in children remains uncertain: 1% is probably conservative, and 10 to 20% is possible. Screening studies of first-degree relatives of individuals with VUR identifies reflux in one third to one half of siblings and 65% of offspring. Futhermore, there is a high concordance of primary VUR in identical twins, and families have been identified with multiple generations affected by primary VUR and RN. Collectively, these studies suggest that there is a substantial genetic component to VUR.

The first genome-wide linkage analysis for VUR, based on seven kindreds, provided preliminary evidence for a locus on chromosome 1 and also for genetic heterogeneity. In this study, multipoint parametric and nonparametric linkage analysis was undertaken; however, one of the markers defining the interval on chromosome 1, GATA176C01, was subsequently found to be on chromosome 2 (Ensembl release 55, July 2009), so this localization should be treated with caution. Subsequent studies using similar kindreds have supported the notion that the condition is genetically heterogeneous. In the largest linkage study of VUR before this report, Kelly et al. performed a linkage genome scan of 609 individuals (283 affected individuals in 129 families) and detected six to seven regions with suggestive evidence of linkage, one of which at chromosome 2q37 attained genome-wide significance when analyzed in a phenotypically derived subset of the data. The high incidence in offspring of affected individuals and the large number of pedigrees consistent with autosomal dominant inheritance, albeit with reduced penetrance, is in keeping with a dominant model; however, recently, a locus was identified on 12p11-q13, albeit with reduced penetrance, is in keeping with a dominant model under a recessive model.

Here we report on linkage and association analysis in affected sibling pairs from two populations. We used the Affymetrix NspI array to generate genome-wide data, adding in haplotype-tagging single-nucleotide polymorphisms (SNPs) to obtain full coverage for six candidate genes: AGTR2, HNF1B, PAX2, RET, ROBO2, and UPK3A.

RESULTS

Linkage Analysis

Disease status was initially coded as positive for cases with VUR and/or RN. Supplemental Figures 1 through 4 show the nonparametric linkage analysis results and information content across the genome for the UK (348 cases in 172 families), Slovenian (313 cases in 148 families), and combined (661 cases in 320 families) samples, respectively. Table 1 shows all results reaching a significance threshold of equivalent logarithm of odds (LOD) score >1. The linkage analysis was repeated for the UK and combined data coding disease status as positive only when there was documented VUR (302 UK cases), that is, excluding cases that had RN without documented VUR (for the Slovenian data set, positive disease status for the original 313 cases already corresponded solely to VUR). The linkage results were very similar to those found using the wider phenotype definition (data not shown); however, the peaks on chromosome 6 in the UK data set and on chromosomes 10 and 11 in the combined data set all increased in significance (LOD = 2.35, \( P = 0.0005 \) at rs863820 on chromosome 6; LOD = 2.32, \( P = 0.0005 \) at rs7904367 at 160.38 cM on chromosome 10; and LOD = 2.27, \( P = 0.0006 \) at rs9733150 on chromosome 11). The linkage analysis was also repeated for a separate subset of the UK families (132 families, 212 cases) in which disease status corresponded to being positive for RN. Only one peak with LOD > 2 was observed (LOD = 2.02, \( P = 0.001 \), at rs1860436 on chromosome 12.

We also carried out parametric linkage analysis (allowing for heterogeneity) under fully penetrant recessive and dominant models. The results (with disease coded as positive for cases with VUR and/or RN) are shown in Table 2. Several peak heterogeneity LODs (HLODs) occur close to nonparametric linkage analysis peaks, but, in addition, under a recessive model, we found three new peaks: In the Slovenian data, HLOD = 2.72 at rs2162769, and in the combined data, HLOD = 3.02 at rs4893496 and HLOD = 2.87 at rs475188. Also in the UK data, rs928720, which showed weak significance (LOD = 1.46) in nonparametric analysis, showed stronger significance (HLOD = 3.12) in parametric analysis under a recessive model.

Association Analysis

The significant results from the family-based association studies (using SNPs passing medium, stringent, or very stringent quality control [QC] thresholds) are shown in Table 3. Figure 1 shows a Manhattan plot of the genome-wide results from the transmission disequilibrium test (TDT) analysis using a medium-stringency SNP selection criterion. Figures 2 through 4 show quantile-quantile (Q-Q) plots of the TDT statistics obtained in the UK, Slovenian, and combined data sets at different levels of stringency of SNP selection. Results are shown with or without use of a robust clustered sandwich estimator of the variance to correct for any nonindependence between related individuals. In the UK data set, little evidence is seen for association at the very stringent, stringent, or medium thresholds, beyond what is expected from genome-wide testing. The top-ranked SNP at the medium threshold is rs11083021 on chromosome 18 (\( P = 3.06 \times 10^{-5} \)). In the Slovenian data set, at the very stringent criterion, two SNPs on chromosome 5 (rs17144806 and rs4895183) show significance beyond what is expected from genome-wide testing (\( P = 5.81 \times 10^{-7} \) and \( P = 2.55 \times 10^{-6} \), respectively). A third SNP (rs16963279 on chromosome 18) shows significance (\( P = 3.13 \times 10^{-5} \)) at the medium criterion. In the combined data set, at the very stringent SNP selection criterion, little evidence is seen for association,
but two SNPs show significance at the stringent criterion (rs11029158 on chromosome 11 \( P = 1.82 \times 10^{-6} \) and rs1696803 on chromosome 10 \( P = 2.25 \times 10^{-6} \)). These are joined by a third (rs2102860 on chromosome 3; \( P = 7.43 \times 10^{-5} \)) at the medium criterion. Interestingly, this SNP lies only approximately 2.2 Mb from the modest linkage peak (LOD 1.411) seen at rs7635068 in the combined data set.

Family-based association analysis for the UK and combined data was repeated using the subset of 615 from the original 661 cases in which disease status was coded as positive for VUR, rather than positive for VUR and/or RN (see Supplemental Figures 5 and 6). For the UK data, there is little evidence for association (beyond what is expected from genome-wide testing) at any SNP selection criterion, although the top-ranked SNPs remain the same as when using the wider phenotype definition. For the combined data, the same top-ranked SNPs are identified at the stringent and medium criteria as were identified using the wider phenotype definition (rs11029158 on chromosome 11 \( P = 2.7 \times 10^{-7} \), rs1696803 on chromosome 10 \( P = 5.27 \times 10^{-6} \) and rs2102860 on chromosome 3 \( P = 3.5 \times 10^{-7} \)); however, interestingly, with the narrower phenotype definition, these results show increased significance and better separation in the Q-Q plots from the bulk of the results that lie on the straight line with slope 1. We also repeated the association analysis for a separate subset of the UK families (132 families, 212 cases) in which disease status corresponded purely to being positive for RN but found nothing of significance.

Supplemental Figures 7 and 8 show Q-Q plots from the family-based analysis of chromosome X SNPs using UNPHASED. Although none reach genome-wide significance thresholds, it is notable that two SNPs (rs1983167 and rs7881785) passing the very stringent QC threshold show clear departure from the theoretical \( \chi^2 \) distribution (\( P = 5.22 \times 10^{-5} \) and \( P = 6.97 \times 10^{-5} \), respectively, in the combined [UK and Slovenian] data). We found no significant associations with any of the SNPs in AGTR2, HNF1B, PAX2, RET, ROBO2, or UPK3A in either the UK or Slovenian data sets, taking into account the multiple tests (146 tests) performed.

Table 4 shows the significant results from our case-control (SNPTEST and STATA logistic regression) analyses at all SNPs passing the medium QC threshold, using our UK VUR cases together with 2938 Wellcome Trust Case Control Consortium (WTCCC) controls. We used genomic control\(^22\) to adjust our
Table 2. Parametric linkage analysis results (all regions with HLOD > 2 under either a recessive or a dominant model)

<table>
<thead>
<tr>
<th>Population</th>
<th>Chr</th>
<th>Closest Nonparametric Peak</th>
<th>Peak Using Parametric Recessive Model</th>
<th>Peak Using Parametric Dominant Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SNP rs ID cM LOD</td>
<td>SNP rs ID cM LOD α^a HLOD</td>
<td>SNP rs ID cM LOD α^b HLOD</td>
</tr>
<tr>
<td>UK</td>
<td>2</td>
<td>SNP_A-2279980 rs645490 32.90 2.34</td>
<td>SNP_A-1807707 rs4669767 29.73 −1.56 0.40 3.02</td>
<td>SNP_A-2279980 rs645490 32.90 −14.90 0.34 2.37</td>
</tr>
<tr>
<td>UK</td>
<td>6</td>
<td>SNP_A-2289125 rs863820 147.51 1.45</td>
<td>SNP_A-2129101 rs928720 140.19 −0.85 0.41 3.12</td>
<td>SNP_A-2289125 rs863820 147.51 −21.43 0.23 1.08</td>
</tr>
<tr>
<td>Slovenian</td>
<td>5</td>
<td>– – – –</td>
<td>– – – – – –</td>
<td>– – – – – –</td>
</tr>
<tr>
<td>Slovenian</td>
<td>11</td>
<td>SNP_A-2235976 rs10792438 86.98 2.49</td>
<td>SNP_A-4233278 rs2097171 88.17 −0.14 0.53 3.33</td>
<td>SNP_A-2269085 rs535809 32.90 −31.76 0.28 2.61</td>
</tr>
<tr>
<td>Combined</td>
<td>2</td>
<td>SNP_A-2279980 rs645490 32.90 2.32</td>
<td>SNP_A-2060999 rs1018432 33.27 −11.98 0.23 2.17</td>
<td>SNP_A-2279980 rs645490 32.90 −31.76 0.28 2.61</td>
</tr>
<tr>
<td>Combined</td>
<td>6</td>
<td>SNP_A-2289125 rs863820 140.19 1.32</td>
<td>SNP_A-2129101 rs928720 140.19 −0.85 0.41 3.12</td>
<td>SNP_A-2289125 rs863820 140.19 −21.43 0.23 1.08</td>
</tr>
<tr>
<td>Slovenian</td>
<td>11</td>
<td>SNP_A-2235976 rs10792438 86.98 2.49</td>
<td>SNP_A-4233278 rs2097171 88.17 −0.14 0.53 3.33</td>
<td>SNP_A-2269085 rs535809 32.90 −31.76 0.28 2.61</td>
</tr>
<tr>
<td>Combined</td>
<td>2</td>
<td>SNP_A-2279980 rs645490 32.90 2.32</td>
<td>SNP_A-2060999 rs1018432 33.27 −11.98 0.23 2.17</td>
<td>SNP_A-2279980 rs645490 32.90 −31.76 0.28 2.61</td>
</tr>
<tr>
<td>Slovenian</td>
<td>11</td>
<td>SNP_A-2235976 rs10792438 86.98 2.49</td>
<td>SNP_A-4233278 rs2097171 88.17 −0.14 0.53 3.33</td>
<td>SNP_A-2269085 rs535809 32.90 −31.76 0.28 2.61</td>
</tr>
<tr>
<td>Combined</td>
<td>2</td>
<td>SNP_A-2279980 rs645490 32.90 2.32</td>
<td>SNP_A-2060999 rs1018432 33.27 −11.98 0.23 2.17</td>
<td>SNP_A-2279980 rs645490 32.90 −31.76 0.28 2.61</td>
</tr>
<tr>
<td>Combined</td>
<td>6</td>
<td>SNP_A-2289125 rs863820 140.19 1.32</td>
<td>SNP_A-2129101 rs928720 140.19 −0.85 0.41 3.12</td>
<td>SNP_A-2289125 rs863820 140.19 −21.43 0.23 1.08</td>
</tr>
</tbody>
</table>

*aEstimated proportion of linked families.

Table 3. Top family-based association results from TDT (for autosomal loci) and UNPHASED (for chromosome X) analyses

<table>
<thead>
<tr>
<th>Population</th>
<th>SNP Selection Threshold</th>
<th>SNP (Affy ID) rsID Chr</th>
<th>Physical Position</th>
<th>Associated Allele</th>
<th>MAF</th>
<th>T:N^a</th>
<th>Allelic OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>Medium</td>
<td>SNP_A-4206971 rs11083021</td>
<td>18</td>
<td>20052848 45.29 T</td>
<td>0.03</td>
<td>32.4</td>
<td>8.00 (2.83 to 22.68)</td>
<td>3.06 × 10^{-6}</td>
</tr>
<tr>
<td>Slov inian</td>
<td>Very stringent</td>
<td>SNP_A-4219737 rs17144806</td>
<td>5</td>
<td>118244217 124.76 G</td>
<td>0.14</td>
<td>105.47</td>
<td>2.23 (1.58 to 3.15)</td>
<td>0.53</td>
</tr>
<tr>
<td>Slovenian</td>
<td>Medium</td>
<td>SNP_A-2234308 rs16963279</td>
<td>18</td>
<td>28204622 54.84 C</td>
<td>0.07</td>
<td>39.8</td>
<td>4.88 (2.27 to 10.43)</td>
<td>0.13</td>
</tr>
<tr>
<td>Combined</td>
<td>Medium</td>
<td>SNP_A-1946120 rs2102860</td>
<td>3</td>
<td>188593677 197.67 A</td>
<td>0.04</td>
<td>57.15</td>
<td>3.80 (2.15 to 6.71)</td>
<td>8.20 × 10^{-5}</td>
</tr>
<tr>
<td>Combined</td>
<td>Stringent</td>
<td>SNP_A-2266057 rs1696803</td>
<td>10</td>
<td>123426940 146.57 G</td>
<td>0.15</td>
<td>175.97</td>
<td>1.80 (1.41 to 2.31)</td>
<td>0.020</td>
</tr>
<tr>
<td>Combined</td>
<td>Stringent</td>
<td>SNP_A-1835586 rs11029158</td>
<td>11</td>
<td>26062215 42.96 C</td>
<td>0.05</td>
<td>72.25</td>
<td>2.88 (1.83 to 4.54)</td>
<td>0.016</td>
</tr>
<tr>
<td>Combined</td>
<td>Stringent</td>
<td>SNP_A-1998634 rs1983167</td>
<td>X</td>
<td>42604638 66.77 G</td>
<td>0.37</td>
<td>1.46</td>
<td>3.29 (1.29 to 2.09)</td>
<td>0.004</td>
</tr>
<tr>
<td>Combined</td>
<td>Stringent</td>
<td>SNP_A-2182543 rs7881785</td>
<td>X</td>
<td>42610680 66.78 G</td>
<td>0.37</td>
<td>1.62</td>
<td>1.27 (2.27 to 0.70)</td>
<td>0.005</td>
</tr>
</tbody>
</table>

CI, confidence interval; MAF, minor allele frequency; OR, odds ratio.

*aRatio of transmissions to nontransmissions from a heterozygous parent to an affected child.

*bMinor allele is associated with disease (transmitted more often to affected children). In all other cases, the disease associated allele is the major allele.
SNPTEST results for inflation as a result of relatedness between cases (see Supplemental Figure 9 for Q-Q plots and details of adjustment for related cases). We used a robust clustered sandwich estimator of the variance to adjust for relatedness in STATA.\textsuperscript{19,20} Overall, there was close correspondence between the top-ranking results from the STATA logistic regression analysis and the SNPTEST analysis. The most significant results from the STATA analysis were rs1159217 on chromosome 10 ($P = 3.96 \times 10^{-7}$) and rs17306391 on chromosome 11 ($P = 5.67 \times 10^{-7}$) and from the SNPTEST analysis, rs12604993 on chromosome 18 ($P = 2.38 \times 10^{-7}$).

**DISCUSSION**

We presented here the results from a genome-wide linkage and association analysis in primary, nonsyndromic VUR and RN. Parametric linkage analysis, which uses assumptions about the mode of inheritance of the disease, is often used to map diseases to genomic regions harboring the causative gene and has been very successful in identifying genes for Mendelian disorders.\textsuperscript{23–26} Nonparametric linkage analysis seeks genomic re-
regions where pairs or groups of affected relatives share more alleles inherited by descent from their common ancestor(s) than expected by chance, making no assumptions about mode of inheritance. In common with many previous genome scans in complex diseases, our nonparametric linkage analyses found a number of regions showing modest linkage signals that did not generally replicate across study groups (either our own or previous studies\textsuperscript{12,13}). Nonparametric linkage analysis of VUR and/or RN identified one locus (on chromosome 2) with LOD \(>2\) in the UK cases, three loci (on chromosomes 6, 11, and X) with LOD \(>2\) in the Slovenian cases, and three loci (on chromosomes 2, 11, and 21) with LOD \(>2\) in the combined data set, but no peaks with a LOD \(>3\). The peak on chromosome 21 could correspond to a weakly significant result (\(P = 0.006\)) in the study of Kelly et al.\textsuperscript{13} The very weak peak we observed on chromosome 3 in the Slovenian data set could correspond to a similarly weak result (\(P = 0.003\)) in the study.

Figure 3. Q-Q plot of TDT statistics in Slovenian data at different levels of stringency of SNP selection in shown.

Figure 4. Q-Q plot of TDT statistics in combined (UK and Slovenian) data at different levels of stringency of SNP selection is shown.
of Kelly et al. Restricting analysis to those with documented VUR increased the significance to LOD > 2 for the peaks on chromosome 6 in the UK data set and on chromosome 10 in the combined data set. This is of interest because the chromosome 6 and chromosome 10 peaks could potentially correspond to results (P = 0.0003 and P = 0.0005 respectively) found in the study of Kelly et al.

Parametric linkage analysis assuming a recessive disease model provided stronger evidence for linkage at rs484936 (HLOD = 3.02) on chromosome 3 and rs2835104 (HLOD = 3.21) on chromosome 21 in the combined data set, at rs4669767 (HLOD = 3.02) on chromosome 2 and rs928720 on chromosome 6 (HLOD = 3.12) in the UK data set, and at rs2097171 on chromosome 11 (HLOD = 3.33) in the Slovenian data set. Parametric linkage analysis assuming a dominant disease model gave one HLOD > 3, at rs9977677 on chromosome 21. All of these linkage regions contain a large number of genes (between 11 and 34 genes using an “HLOD minus 1” threshold and between 23 and 131 genes using an “HLOD minus 2” threshold). Although HLODs > 3 are encouraging, the fact that we have maximized the LOD score over two models (recessive and dominant) as well as over a heterogeneity parameter (the proportion of families in either of the populations we have studied more often to affected offspring. The TDT provides a test that is robust to population stratification, unlike standard case-control association analyses; however, the advantage of additionally performing case-control analysis in our data is the greater power that can potentially be achieved, first through use of a larger control sample and second by the ascertainment “bias” introduced through the comparison of cases who have a close relative (a sibling) who also has the disease, with standard population control subjects.

A number of loci in the association analyses approached genome-wide significance levels; however, there was relatively little concordance in the results from our two study groups (UK and Slovenian) or between the results from our linkage and association analyses. This is not surprising given the relatively small sample size of the individual population groups, meaning that the power to detect small effects as conferred by most risk alleles in complex diseases will be low, rs11083021 is in intron 3 of OSBPPLA, which encodes oxysterol-binding protein—like 1A, a member of the OSBP family of intracellular lipid receptors. The two chromosome 5 SNPs detected in the Slovenian analysis, rs4895183 and rs17144806, and rs17175928 and rs16963279 are intronic SNPs in genes of unknown function, DTWD2, CI1orf72, and FAM59A, respectively. rs2102860 is in linkage disequilibrium (LD) with RTP4, which encodes a Golgi chaperone that plays a role in movement of μ-δ opioid receptor to the cell surface membrane and may be involved in membrane targeting of other G protein–coupled receptors. rs1983167 and rs7881785 both are in LD with the monoamine oxidase inhibitors MAOA and MAOB. rs12604993 is in LD with TXNL4A (thioredoxin-like 4A). Two SNPs identified in the family-based analysis, rs1696803 and rs11029158, and two identified in the case-control analysis of the UK cases, rs11599217 and rs17306391, are not in LD with transcripts.

We also tested for association with six candidate genes previously implicated in human renal tract development: AGTR2, HNF1B, PAX2, RET, ROBO2, and UPK3A. Mutations have been identified in ROBO2 and HNF1B in patients with urinary tract malformations, but we found no evidence of an asso-

### Table 4. Top-ranking results from case-control analysis performed using either SNPTEST (with or without the ‘proper’ option) or logistic regression in STATA

<table>
<thead>
<tr>
<th>SNP (Affy ID)</th>
<th>rsID</th>
<th>Chromosome</th>
<th>Physical Position cM</th>
<th>Associated Allele</th>
<th>MAF</th>
<th>SNPTEST P (proper)</th>
<th>SNPTEST P (without proper)</th>
<th>STATA P</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP_A-1903023</td>
<td>rs11599217</td>
<td>10</td>
<td>128556954</td>
<td>158.28</td>
<td>A*</td>
<td>0.39</td>
<td>2.98 × 10^-5</td>
<td>1.95 × 10^-6</td>
</tr>
<tr>
<td>SNP_A-2235245</td>
<td>rs17306391</td>
<td>11</td>
<td>23147948</td>
<td>38.74</td>
<td>G*</td>
<td>0.03</td>
<td>0.000051</td>
<td>1.17 × 10^-6</td>
</tr>
<tr>
<td>SNP_A-2243575</td>
<td>rs12604993</td>
<td>18</td>
<td>75866208</td>
<td>121.56</td>
<td>G*</td>
<td>0.19</td>
<td>2.38 × 10^-7</td>
<td>1.96 × 10^-6</td>
</tr>
</tbody>
</table>

SNPTEST results are adjusted via genomic control to account for relatedness between individuals, whereas STATA results are adjusted through use of a robust clustered sandwich estimator of the variance. MAF, minor allele frequency.

*Minor allele is associated with disease.

**Major allele is associated with disease.
association between common alleles in these genes and VUR in this study. Yim et al. reported association between an *AGTR2* intronic variant and diverse kidney malformations, although this was not replicated in two studies. Yang et al. presented data for association with between a *RET* polymorphism (p.Gly691Ser) and VUR, but this was not replicated in an Irish cohort. Jiang et al. reported a weak association between a *UPK3A* missense polymorphism and VUR. We have not detected significant association with any of these genes and VUR in the UK, Slovenian, or combined VUR patient groups.

The final size of our study (320 families, 661 affected individuals, 140,484 SNPs) is modest in size and genome coverage is incomplete; nevertheless, it is by far the largest genetic study of VUR to date. Given the small sample size of our collection, our results should be interpreted with caution. Our genome-wide association analyses are perhaps best considered as exploratory, the findings from which require replication in larger cohorts. An obvious question regarding the association analysis is what coverage of the genome was achieved. The Affymetrix 500K SNP array set (consisting of two arrays, Sty and Nsp) has been previously estimated to provide 65% coverage of the genome at $R^2 = 0.8$. This leaves a 35% chance that none of the SNPs genotyped reaches this level of correlation with a causal variant. Our study (conceived originally as a linkage study) used only a single array from the 500K set (the Nsp array), which approximately halves the number of SNPs genotyped. QC procedures reduced the number of SNPs still further, reducing our coverage to levels that are probably closer to the 31% provided by the Affymetrix 11K array. Increasing sample size and conducting a whole-genome association scan with a much denser array, taking advantage of recent improvements in genotyping quality and genotype calling, is an approach that has achieved recent success in other complex genetic diseases, is the natural next step in testing whether common variants contribute to VUR.

**CONCISE METHODS**

**Sample Collection**
The study, which adhered to the Declaration of Helsinki, was approved by UK Research Ethics Committees and the Slovenian National Ethics Committee; informed consent was obtained before sample collection. In the United Kingdom, families were referred by primary physicians from the UK VUR Study Group (listed in acknowledgments). The inclusion criteria for UK families were an index case with VUR diagnosed using x-ray cystography or radionuclide cystography, together with an affected sibling with radiologically proven VUR and/or RN demonstrated on DMSA scanning. Blood samples were collected from affected siblings and their parents, and DNA was extracted by standard procedures. The UK collection comprises 189 index cases with 219 affected siblings (161 sibling pairs, 26 families with three affected children, and two families with four affected children). A total of 120 of the index cases had both VUR and RN, 60 had VUR only, and information on RN was not available for the remaining nine index cases; 77 of the siblings had both VUR and RN, 78 had VUR only, 40 had RN only, and information on RN was not available in the remaining 24 siblings with VUR.

Slovenian patients were identified from the database of children who were referred to the Department of Pediatric Nephrology, University Medical Centre Ljubljana. Suitable families were identified by screening the database for children who had VUR and for whom the reason for investigation was a sibling with VUR. VUR was diagnosed by voiding urosonography, radionuclide cystography, or, in a few cases, x-ray cystography. Blood samples were collected from affected siblings and their parents, and DNA was extracted by standard procedures. The Slovenian collection comprises 149 index cases with 169 affected siblings (133 sibling pairs, 13 families with three affected children, two families with four affected children, and one family with five affected children).

All of the cases we collected had primary VUR (i.e., there were no cases with anatomic or neurogenic bladder outflow obstruction), and we excluded families in which the index case or sibling had additional structural defects in the urinary tract. In both the UK and Slovenian collections, parental disease status was coded as unknown. Parental DNA was genotyped in 612 parents from the 320 families. Three unaffected siblings (from the UK collection) were also genotyped and included in the linkage analysis.

**DNA Analysis**
Genotyping of the samples was carried out by the company Geneservice, using the Affymetrix 262,264 SNP NspI array. Genotypes at the 262,264 SNPs were assigned (“called”) from the raw intensity data using the CHIAMO algorithm. For the six candidate genes, we used the Tagger option in the program Haploview to identify SNPs to provide coverage at $R^2 = 0.8$. This resulted in 101 SNPs for *ROBO2*, 12 for *RET*, 10 for *PAX2*, 17 for *HNF1B*, five for *UPK3A*, and three for *AGTR2* (Supplemental Table 1), and these were typed by Sequenom.

**Quality Control**
Stringent QC checks were used to ensure the accuracy of the final genotype data and pedigree information (details in Supplemental information). The final number of samples remaining after all of the sample QC checks had been carried out was 1282 (comprising 692 UK samples and 590 Slovenian samples) in 320 families (172 UK families and 148 Slovenian families) of an original 1398 samples genotyped. We used these samples to re-perform QC measures on the SNPs, to choose SNPs with the most reliable genotype calls for the final analysis. We selected SNPs at five different levels of stringency, as shown in Supplemental Table 2.

**Linkage Analysis**
We used SNPs passing the very stringent QC threshold to perform multipoint nonparametric (model-free) linkage analysis across the genome. For reasons of computational efficiency, we thinned our set of SNPs to use a single SNP—that with the highest heterozygosity—in each 1-cM window. Examination of the resulting information content plots (Supplemental Figure 4) indicates that this thinned set of SNPs provides adequate linkage information. We used the programs MERLIN and MINX to calculate information content and test for linkage...
using a multipoint “equivalent LOD score” to the Kong and Cox exponential model likelihood-based allele-sharing test.\textsuperscript{15} We also used MERLIN and MINX to perform parametric linkage analysis allowing for heterogeneity (an “HLOD” analysis), assuming a disease allele frequency of 0.01, under both recessive (penetrances 0.01, 0.01, and 0.99) and dominant (penetrances 0.01, 0.99, and 0.99) models.\textsuperscript{27}

**Association Analysis**

We used the TDT\textsuperscript{52} at each SNP passing our various QC filters. We calculated the TDT using the R package DGCGenetics (1) assuming each case/parent trio was an independent unit and (2) allowing for nonindependence between related trios (e.g., affected sibling pairs) through use of a robust clustered sandwich estimator of the variance.\textsuperscript{19,20} Significance was assessed through examination of Q-Q plots,\textsuperscript{18} which is broadly equivalent to use of a Bonferroni correction to assess the overall significance of a given result in light of the multiple tests performed.

The candidate gene SNPs were analyzed by TDT. Because analysis of these genes was hypothesis driven, we did not apply the same multiple correction factor as already discussed but instead corrected for the 146 independent tests.

For the UK cases, we additionally used the affected offspring (cases) together with a sample of 2938 control subjects (genotyped at the same SNPs by the WTCCC\textsuperscript{18}) to perform case-control analysis. We used the program SNPTTEST\textsuperscript{18} to perform frequentist case-control association tests at each SNP passing our medium QC filters, first using the default options and then additionally allowing for genotype uncertainty via the ‘-proper’ option within SNPTTEST. We incorporated as covariates in the analysis the first six principal components from a principal components analysis performed using the ‘smartpca’ routine within the EIGENSOFT package,\textsuperscript{53} using a set of 45,459 markers chosen to be in low LD with one another ($R^2 < 0.2$) and using only unrelated individuals (the WTCCC control subjects and a single case from each VUR sibship) to infer the eigenvectors, onto which the remaining individuals were then projected. This approach has been previously proposed to adjust for population stratification.\textsuperscript{53} Logistic regression (incorporating the same six covariates) in STATA was also performed at each SNP (1) assuming that each case/parent trio was independent and (2) allowing for nonindependence between related cases through use of a robust clustered sandwich estimator of the variance.\textsuperscript{19,20} We used genomic control\textsuperscript{22} to adjust our SNPTEST results for inflation as a result of relatedness between cases.

We analyzed SNPs on chromosome X separately from the autosomal SNPs, using, for the family data, a likelihood-based analysis as implemented in the program UNPHASED\textsuperscript{23} and, for the case/control data, logistic regression analysis including the first six principal components as covariates as described already, with robust clustered variance estimates. The rationale for treating chromosome X SNPs separately was that many of the methods and programs described already for autosomal analysis are not directly applicable to X-linked loci.\textsuperscript{54}

Visual assessment of the cluster plots\textsuperscript{18} on which genotype calls were based (see supplemental information) was performed for SNPs showing significant association with disease. All except one SNP (rs16963279 on chromosome 18 from Table 3) showed reasonable separation among the three genotype clusters. rs16963279 showed only two genotype clusters, probably because the minor allele frequency is sufficiently low that no homozygotes for the minor allele were observed in our samples.

**ACKNOWLEDGMENTS**

This project was supported by the British Association for Pediatric Nephrology (BAPN), the Medical Research Council (MRC) (Grant reference G0600040), and the Wellcome Trust (Grant references 066647, 070327, and 074524). Members of the UK VUR Study Group were Royal Victoria Infirmary, Newcastle upon Tyne: M. Coulthard, H. Lambert, E. Hunter, M. Kier, N. Moghal, M. Ognanovic, S. Vernon; Cumberland Infirmary: J. Storr; West Cumberland Infirmary: J. Jackson; University Hospital of North Tees: I. Verber; James Cook University Hospital, Middlesbrough: S. Sinha; Leeds Teaching Hospitals NHS Trust: M. Fitzpatrick, S. Feather; Royal Manchester’s Children’s Hospital: M. Lewis, N. Webb, M. Bradbury, N. Plant, R. Postlethwaite, D.J. O’Donoghue; Royal Liverpool Children’s Hospital: D. Hughes, C. Jones, B. Judd; Royal Belfast Hospital for Sick Children: M. Savage, M. O’Connor, M. Convery; Burnley General Hospital: J. Iqbal; Royal Hospital for Sick Children, Glasgow: H. Maxwell, J. Beattie, Royal Hospital for Sick Children, Edinburgh: S. Taheri; Great Ormond Street Hospital, London: P. Cuckow, S. Marks, L. Rees, R. Trompeter, K. Tullius, W. Van’t Hoff, D. Wilcox, A. Woolf; Evelina Children’s Hospital, London: G. Haycock, C. Reid, S. Rigdon; Nottingham City Hospital: A. Watson; Birmingham Children’s Hospital: S. Hulton, D. Milford, S. Stephens, C.M. Taylor; Bristol Royal Hospital for Sick Children: J. Dudley, C. Inward, M. McGraw, J. Tizard; University Hospital of Wales: K. Verriker-Jones; St. Mary’s Hospital, Portsmouth: J. Scanlan; Leicester Royal Infirmary: P. Houtman; Addenbrookes Hospital, Cambridge: R. Sandford; Queen Elizabeth the Queen Mother Hospital, Margate: E. Rfidah; University College Hospital, London: D. Hodes, A. Kilby; Walsgrave Hospital, Coventry: N. Coad; Derriford Hospital, Plymouth: R. Jones; Gloucester Royal Hospital: L. Jadresic; Northampton General Hospital: N. Griffin; Queen Elizabeth Hospital, Kings Lynn: J. Dossetor, A. Hughes; Whittington Hospital, London: M. Jaswon.

We thank Hin-Tak Leung, Jonathan Marchini, and Chris Spencer for assistance with use of software developed by the WTCCC. We thank and acknowledge the oversight provided to the collectors through the VUR Development Group, led by Kidney Research UK, which included representation from the wider research community and the BAPN.

**DISCLOSURES**

None.

**REFERENCES**

2. Yeung CK, Godley ML, Dhillon HK, Gordon I, Duffy PG, Ransley PG:


20. Welcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678, 2007


---

Supplemental information for this article is available online at http://www.jasn.org/.