

## Supplemental Material Table of Contents

**Supplemental Figure 1:** PCA1-4 in Cases and Controls in the UK biobank. Available at UK Biobank data field 22009.

**Supplemental Figure 2:** Q-Q Plot for Centrally HRC-imputed Hematuria GWAS. The Q-Q plot is a graphical representation of the deviation of the observed P values from the null hypothesis stratified by quartiles of minor allele frequencies. Figure generated using summary statistics from <https://pheweb.org/UKB-SAIGE/pheno/593>.

**Supplemental Table 1.** Descriptive characteristics of UK Biobank white British subset used in the hematuria analyses.

**Supplemental Table 2.** Association of 3 independent *COL4A4-COL4A3* variants with other kidney-disease relevant or Alport syndrome traits using TOPMed imputed UK Biobank dataset. Source is pheweb.org.

**Supplemental Table 3.** GWAS via SAIGE using *HLA* haplotypes provided by UK Biobank on white British for ICD-based case-control hematuria. After conditioning on *HLA-B\*0801*, which was the strongest association (row 59, red), no other associations were statistically significant (highlighted in blue) including IgA nephropathy haplotypes reported by Kiryluk K et al Nature Genetics 2014 (PMID:25305753) shown in rows 275 (*DQB1\*0201*, yellow), 277 (*DQB1\*0301*, yellow), 294 (*DQA1\*0102*, yellow) and 295 (*DQA1\*0103*, yellow).

**Supplemental Table 4.** Novel hematuria loci in UK Biobank are replicated in deCODE dataset, though association with variant, nearest gene *PDPN* is nominal. Results for phenotype defined as  $\geq 2+$  hematuria or  $\geq 1+$  hematuria are both shown for each of the novel loci.

**Supplemental Table 5.** Cell and tissue specific expression of novel loci in hematuria GWAS. Associated phenotypes from pheweb.org also shown. LH=loop of Henle, DCT=distal convoluted tubule, CNT=connecting tubule, PC=principal cell, IC-A=intercalated cell A, EC=endothelial cell

**Supplemental Table 6.** Estimated pairwise linkage disequilibrium between UK Biobank imputed haplotypes *HLA-B\*0801*, *HLA-DQB1\*0201* and *HLA-DQB1\*0301* for the white British hematuria cases and controls.

**Supplemental Table 7.** *HLA* haplotype frequencies.

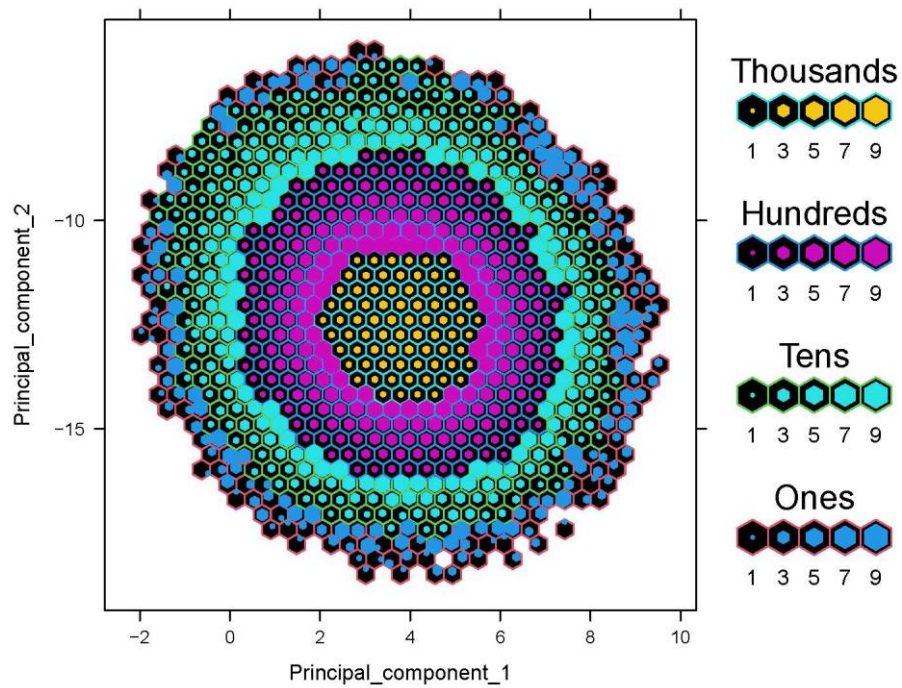
**Supplemental Table 8.** Estimated haplotype frequencies for *HLA-B\*0801* and *HLA-DQB1\*0201* in white British in UK Biobank.

**Supplemental Table 9.** Estimated haplotype frequencies for *HLA-B\*0801* and *HLA-DQB1\*0301* in white British in UK Biobank.

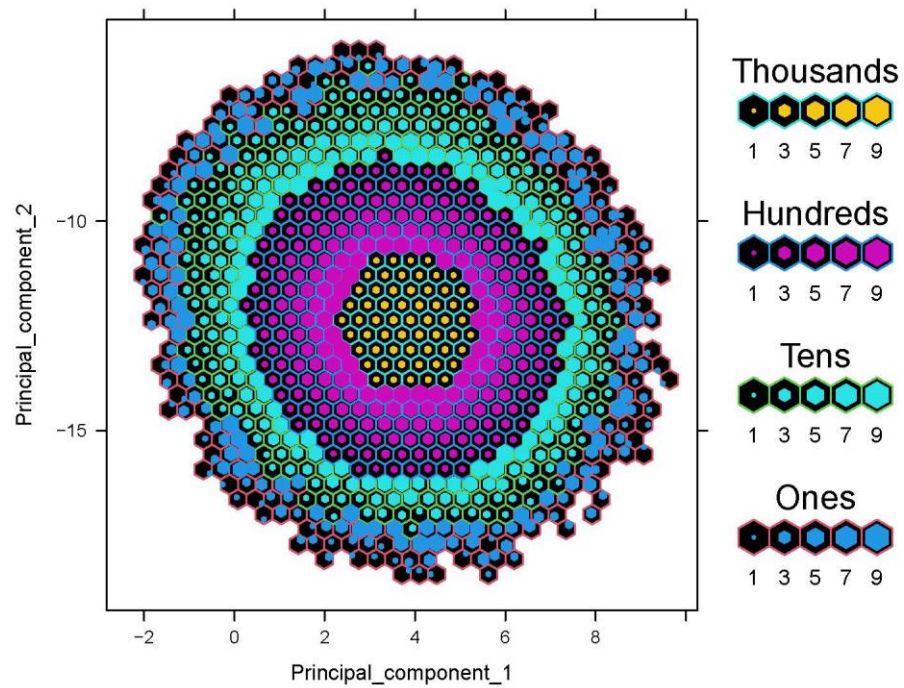
**Supplemental Table 10.** Estimated haplotype frequencies for *HLA-DQB1\*0201* and *HLA-DQB1\*0301* in white British in UK Biobank.

**Supplemental Figure 1:** PCA1-4 in Cases and Controls in the UK biobank. Available at UK Biobank data field 22009.

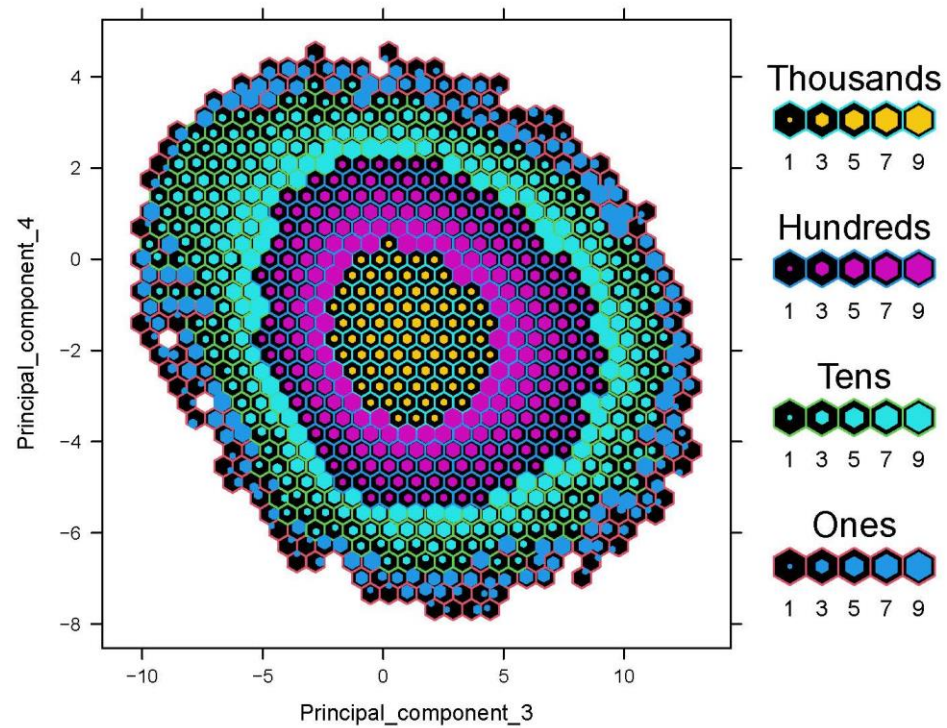
Controls



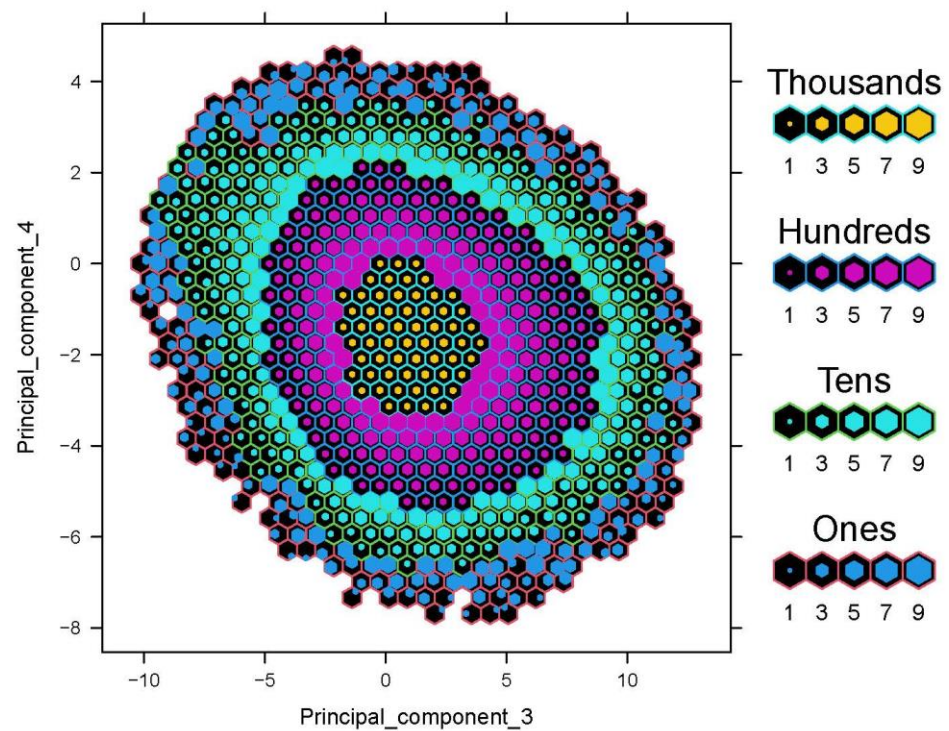
Cases



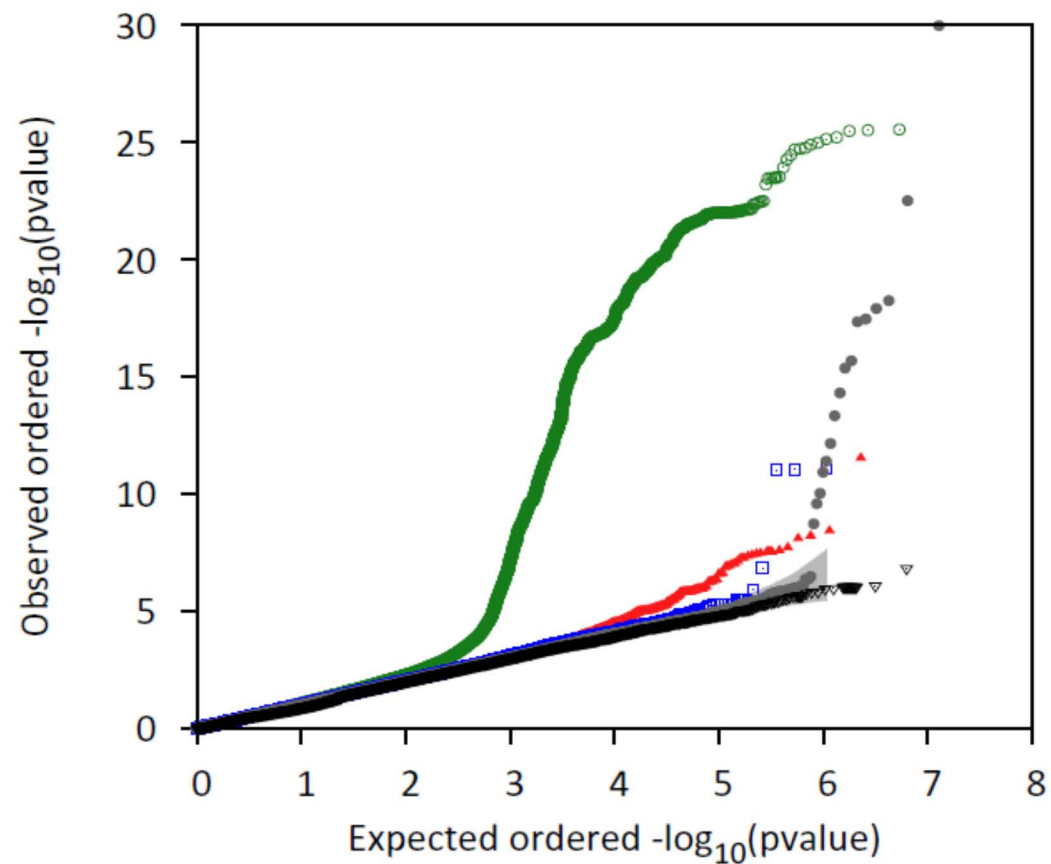
Controls



Cases



**Supplemental Figure 2:** Q-Q Plot for Centrally HRC-imputed Hematuria GWAS. The Q-Q plot is a graphical representation of the deviation of the observed P values from the null hypothesis stratified by quartiles of minor allele frequencies. Figure generated using summary statistics from <https://pheweb.org/UKB-SAIGE/pheno/593>.



MAF [.05,.5] (5432396)	○
MAF [.01,.05] (2310010)	▲
MAF [.005,.01] (1074851)	□
MAF [.0001,.005] (13086692)	●
MAF (0,.0001) (6328746)	▼

**Supplemental Table 1:** Descriptive characteristics of UK biobank white British subset used in the hematuria analyses.

	Cases	Controls	Cases with non-missing data	Controls with non-missing data	T-test p-value
Total (female+male)	16,866 (7,074+9,292)	391,420 (213,407+178,013)	--	--	< 2.2E-16 (chi-square)
Age at baseline in years, mean±SD	59.7±7.2	56.8±8.0	16,866	391,420	< 2.2E-16
Number of individuals on the 2 arrays	UK BiLEVE: 2043 UK Biobank Axiom: 14,823	UK BiLEVE: 42,542 UK Biobank Axiom: 348,878	16,866	391,420	3.90E-07 (chi-square)
PC1, mean±SD	-12.36±1.61	-12.36±1.61	16,866	16,866	0.75
PC2, mean±SD	3.78±1.50	3.78±1.50	16,866	16,866	0.99
Pulse rate, automated reading at baseline in bpm, mean±SD	70.0±12.1	69.5±11.7	15,731	365,689	7.1E-07
Systolic blood pressure, automated reading at baseline in mmHg, mean±SD	142.7±19.8	140.1±19.6	15,731	365,682	< 2.2E-16
Diastolic blood pressure, automated reading at baseline in mmHg, mean±SD	82.7±10.8	82.2±10.6	15,731	365,682	1.115E-08
Taking blood pressure medication at baseline	2336 (1010 females + 1326 males)	38,892 (21,438 females + 17,454 males)	16,842 (7061 females + 9781 males)	390,756 (213,030 females + 177,726 males)	< 2.2E-16 (chi-square)
Standing height at baseline in cm, mean±SD	169.8±9.26	168.7±9.25	16,798	390,632	< 2.2E-16
Weight at baseline in Kg, mean±SD	81.24±16.7	78.15±15.8	16,781	390,371	< 2.2E-16
BMI at baseline, mean±SD	28.08±4.93	27.38±9.26	16,773	390,233	< 2.2E-16
Urinary Albumin-Creatinine Ratio in mg/mmol*, Median (1 <sup>st</sup> -3 <sup>rd</sup> percentile)	1.19 (0.73-2.13)	1.10 (0.69-1.85)	16,399	380,240	< 2.2E-16 (Mann-Whitney U test)
eGFR in mL/min/1.72m <sup>2</sup> **, mean±SD	87.39±14.6	90.67±13.2	16,123	372,966	< 2.2E-16

eGFR computed as 100/serum cystatin C (mg/L) at baseline, mean±SD	107.53±18.5	113.3± 17.6	16,122	373,134	< 2.2E-16
Hearing aid user (yes; no)	809; 9,617	11,876; 223,977	10,426	235,853	< 2.2E-16 (chi-square)
Hearing difficulty/problems via touchscreen (yes; no)	5,205; 11,014	97,706; 278,260	16,219	375,966	< 2.2E-16 (chi-square)



**Supplemental Table 2:** Association of 3 independent *COL4A4-COL4A3* variants with other renally relevant or Alport syndrome traits using TOPMed imputed UK Biobank dataset. Source is pheweb.org. Effect size is to alternate allele.

Gene, SNP	Phenotype	P-value	Beta (SE)	Number of samples (cases/controls)	
<i>COL4A4</i> p.Ser969X rs35138315	Hematuria	8.8E-49	4.5 (0.31)	16235/378356	
	Essential hypertension	6.8E-7	0.61 (0.12)	77465/328796	
	Hypertension	8.0E-7	0.61 (0.12)	77714/328796	
	Other symptoms/disorders of the urinary system	8.9E-5	3.6 (0.92)	1387/396050	
	Retinal detachment with retinal defect	9.6E-5	3.6 (0.92)	1387/396050	
	Nephritis and nephropathy with pathological lesion	1.3E-4	11 (2.9)	252/395835	
	Chronic glomerulonephritis, NOS	1.4E-4	4.6 (1.2)	877/395835	
	Nephritis; nephrosis; renal sclerosis	2.0E-4	3.2 (0.85)	1531/395835	
	Glomerulonephritis	4.3E-4	3.7 (1.1)	1052/395835	
	Proteinuria	5.6E-4	7.4 (2.1)	337/406470	
	Myopia	7.5E-4	3.3 (0.97)	1249/404788	
	Chronic renal failure	8.4E-4	2.0 (0.60)	2649/395832	
	Retinal detachments and defects	9.4E-4	1.8 (0.55)	3244/396051	
	<i>COL4A4</i> intron 25 variant rs58261427	Hematuria	1.2E-9	0.086 (0.014)	16235/378356
		Hearing loss	7.6E-3	0.072 (0.027)	4192/402829
	Conductive hearing loss	9.3E-3	0.21 (0.082)	449/402828	

	Chronic glomerulonephritis, NOS	9.9E-3	-0.15 (0.059)	877/395835
	Glomerulonephritis	1.1E-2	-0.14 (0.054)	1053/395835
	Chronic renal failure	1.1E-2	-0.087 (0.034)	2649/395832
	Sensorineural hearing loss	1.5E-2	0.20 (0.081)	455/402829
<hr/>				
<i>COL4A3</i> p.Gly695Arg rs200287952	Hematuria	1.7E-8	3.9 (0.68)	16235/378356
	Type 1 diabetes with ophthalmic manifestations	1.2 E-3	18 (5.7)	473/387081
	Type 2 diabetes with ophthalmic manifestations	1.6E-3	10 (3.2)	986/387082
	Diabetic retinopathy	3.6E-3	7.5 (2.6)	1330/395155
	Other hypertensive complications	8.0E-3	28 (11)	144/328788
	Benign neoplasm of kidney and other urinary organs	9.2E-3	24 (9.3)	173/369609
	Small kidney	1.1E-2	23 (9.0)	179/395834
	Preeclampsia and eclampsia	1.2E-2	19 (76)	306 (218980)

**Supplemental Table 5.** Cell and tissue specific expression of novel loci in hematuria GWAS. Associated phenotypes from pheweb also shown. LH=loop of Henle, DCT=distal convoluted tubule, CNT=connecting tubule, PC=principal cell, IC-A=intercalated cell A, EC=endothelial cell

Nearest gene and SNP	snRNAseq in healthy human adult kidney (4,524 nuclei; humphreyslab.com)	GTex V8 Median TPM in kidney	PheWas <a href="https://pheweb.org/UKB-TOPMed/">https://pheweb.org/UKB-TOPMed/</a>
<i>PDPN</i> , rs2885134	podocyte	2.193 (cortex), 1.748 (medulla)	Hematuria, p=4.9E-9, effect size 0.21 (0.035)
<i>SORL1</i> , rs618048	LH, DCT, CNT, PC, IC-A	8.9 (cortex), 14.46 (medulla)	-
<i>PLLPL</i> , rs948705	PT (S3), IC-A, IC-B	9.247 (cortex), 9.425 (medulla)	-
<i>CCDC97</i> , <i>TGFBI</i> , rs56254331	EC, IC-A ( <i>TGFBI</i> )	<i>CCDC97</i> : 9.45 (cortex), 11.66 (medulla); <i>TGFBI</i> : 32.47 (cortex), 46.64 (medulla)	Hematuria, p=8.1E-16, effect size -0.12 (0.15) other disorders of the kidney and ureters, p=1.6E-3, effect size=-0.10 (0.33)

**Supplemental Table 6.** Estimated pairwise linkage disequilibrium between UK Biobank imputed haplotypes *HLA-B*\*0801, *HLA-DQB1* \*0201 and *HLA-DQB1* \*0301 for the white British hematuria cases and controls.

Haplotype pair	r <sup>2</sup>	D'
<i>HLA-B</i> *0801 & <i>HLA-DQB1</i> *0201	0.47	0.69
<i>HLA-B</i> *0801 & <i>HLA-DQB1</i> *0301	0.01	0.28
<i>HLA-DQB1</i> *0201 & <i>HLA-DQB1</i> *0301	0.02	0.34

**Supplemental Table 7.** *HLA* haplotype frequencies.

Haplotype	Frequency
<i>HLA-B</i> *0801 & <i>HLA-DQB1</i> *0201	18%
<i>HLA-B</i> *0801 & <i>HLA-DQB1</i> *0301	5.2%
<i>HLA-DQB1</i> *0201 & <i>HLA-DQB1</i> *0301	5.0%

**Supplemental Table 8.** Estimated haplotype frequencies for *HLA-B*\*0801 and *HLA-DQB1*\*0201 in white British in UK Biobank.

<i>HLA-B</i> *0801	<i>HLA-DQB1</i> *0201	Observed Frequency	Expected Frequency under linkage equilibrium
<i>HLA-B</i> *0801	<i>HLA-DQB1</i> *0201	0.188	0.061
not	<i>HLA-DQB1</i> *0201	0.063	0.190
<i>HLA-B</i> *0801	not	0.056	0.183
not	not	0.693	0.566

**Supplemental Table 9.** Estimated haplotype frequencies for *HLA-B*\*0801 and *HLA-DQB1*\*0301 in white British in UK Biobank.

<i>HLA-B</i> *0801	<i>HLA-DQB1</i> *0301	Observed Frequency	Expected Frequency under linkage equilibrium
<i>HLA-B</i> *0801	<i>HLA-DQB1</i> *0301	0.052	0.072
not	<i>HLA-DQB1</i> *0301	0.231	0.211
<i>HLA-B</i> *0801	not	0.203	0.183
not	not	0.513	0.533

**Supplemental Table 10.** Estimated haplotype frequencies for *HLA-DQB1\*0201* and *HLA-DQB1\*0301* in white British in UK Biobank.

<i>HLA-DQB1*0201</i>	<i>HLA-DQB1*0301</i>	Observed Frequency	Expected Frequency under linkage equilibrium
<i>HLA-DQB1*0201</i>	<i>HLA-DQB1*0301</i>	0.050	0.075
not	<i>HLA-DQB1*0301</i>	0.236	0.210
<i>HLA-DQB1*0201</i>	not	0.214	0.188
not	not	0.501	0.526