





OPEN ACCESS

Original research

Recurrent 17q12 microduplications contribute to renal disease but not diabetes

Stuart Cannon ¹, Rhian Clissold,² Kittiya Sukcharoen,¹ Marcus Tuke,¹ Gareth Hawkes,¹ Robin N Beaumont,¹ Andrew R Wood,¹ Mark Gilchrist ¹, Andrew T Hattersley,¹ Richard A Oram,¹ Kashyap Patel,¹ Caroline Wright,¹ Michael N Weedon¹

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/jmg-2022-108615>).

¹Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK

²Exeter Kidney Unit, Royal Devon University Healthcare NHS Foundation Trust, Exeter, UK

Correspondence to

Mr Stuart Cannon, University of Exeter Medical School, Exeter, UK; S.Cannon@exeter.ac.uk

Received 5 April 2022

Accepted 3 September 2022

ABSTRACT

Background 17q12 microdeletion and microduplication syndromes present as overlapping, multisystem disorders. We assessed the disease phenotypes of individuals with 17q12 CNV in a population-based cohort.

Methods We investigated 17q12 CNV using microarray data from 450 993 individuals in the UK Biobank and calculated disease status associations for diabetes, liver and renal function, neurological and psychiatric traits.

Results We identified 11 17q12 microdeletions and 106 microduplications. Microdeletions were strongly associated with diabetes ($p=2\times 10^{-7}$) but microduplications were not. Estimated glomerular filtration rate (eGFR mL/min/1.73 m²) was consistently lower in individuals with microdeletions ($p=3\times 10^{-12}$) and microduplications ($p=6\times 10^{-25}$). Similarly, eGFR <60, including end-stage renal disease, was associated with microdeletions ($p=2\times 10^{-9}$, $p<0.003$) and microduplications ($p=1\times 10^{-9}$, $p=0.009$), respectively, highlighting sometimes substantially reduced renal function in each. Microduplications were associated with decreased fluid intelligence ($p=3\times 10^{-4}$). SNP association analysis in the 17q12 region implicated changes to *HNF1B* as causing decreased eGFR (NC_000017.11:g.37741642T>G, rs12601991, $p=4\times 10^{-21}$) and diabetes (NC_000017.11:g.37741165C>T, rs7501939, $p=6\times 10^{-17}$). A second locus within the region was also associated with fluid intelligence (NC_000017.11:g.36593168T>C, rs1005552, $p=6\times 10^{-9}$) and decreased eGFR (NC_000017.11:g.36558947T>C, rs12150665, $p=4\times 10^{-15}$).

Conclusion We demonstrate 17q12 microdeletions but not microduplications are associated with diabetes in a population-based cohort, likely caused by *HNF1B* haploinsufficiency. We show that both 17q12 microdeletions and microduplications are associated with renal disease, and multiple genes within the region likely contribute to renal and neurocognitive phenotypes.

INTRODUCTION

Recurrent microdeletions (OMIM: 614527) and microduplications (OMIM: 614526) of ~1.4 Mb region of 17q12 have each been associated with a range of conditions including diabetes, renal disease and intellectual disability (see Rasmussen *et al*¹ for

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ 17q12 microdeletions (OMIM: 614527) and microduplications (OMIM: 614526) are known to cause complex, multisystem syndromes, including diabetes and renal disease.
- ⇒ Disease phenotypes in microdeletions are predominantly attributed to the haploinsufficiency of the 17q12 gene, *HNF1B*.

WHAT THIS STUDY ADDS

- ⇒ Here we show that 17q12 microduplications as well as microdeletions contribute to renal disease in a population cohort.
- ⇒ We also show that genes other than *HNF1B* in the 17q12 region are implicated in renal disease caused by microduplications.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ This study highlights the need to monitor duplication patients for renal disease, but also how further research to delineate the genes driving the complex syndromes is required.

review). Despite incomplete penetrance and variable expressivity,² some phenotypes present more consistently with microdeletions and others with microduplications.^{3–5} The microdeletion syndrome can include developmental kidney disease, early-onset diabetes mellitus, pancreatic hypoplasia, genital tract malformations, abnormal liver function tests, and neurodevelopmental disorders such as autism spectrum disorder and attention deficit hyperactivity disorder.^{2,6} The 17q12 microduplication is associated with a more variable phenotype that can include cognitive impairment, speech and motor developmental delay, brain anomalies, dysmorphic facial features, behavioural abnormalities, oesophageal atresia, renal anomalies and epilepsy.^{3,7–9} Renal function in 17q12 microduplications has not been characterised in a population cohort.

The 17q12 CNV region encompasses 15 protein-coding genes, including *HNF1B*, bounded by segmental microduplications. The association of the 17q12 deletion with diabetes and decreased kidney function is known to be caused by haploinsufficiency of *HNF1B*.^{10,11} Protein-truncating



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY. Published by BMJ.

To cite: Cannon S, Clissold R, Sukcharoen K, *et al*. *J Med Genet* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jmedgenet-2022-108615

variants (PTVs) in *HNF1B* cause a similar phenotype^{6 12 13}; however, *HNF1B* PTVs are not associated with neurodevelopmental disorders.¹⁴ It is unknown whether *HNF1B* triplosensitivity contributes to the phenotypes observed in individuals with 17q12 microduplications or whether other genes in the region also play a role in the deletion or microduplication phenotypes.

Individuals with 17q12 microdeletions and microduplications are often identified through clinical referral and present with distinct phenotypes.¹⁵ However, large data sets are required to determine the functional associations of rare CNVs with variable penetrance and expressivity in the population.¹⁶ The phenotypes of individuals with these variants when identified incidentally from the population have not been studied in depth. The UK Biobank (UKB), a population-based cohort of ~500 000 individuals, offers a unique opportunity to characterise the 17q12 locus in a population setting and to investigate the differences between deletion and microduplication phenotypes. In this study we assessed diabetes plus renal, liver and neurological phenotypes of individuals in the UKB with recurrent 17q12 microdeletions and microduplications. We demonstrate that microduplications are a cause of renal disease but not diabetes and provide evidence that genes other than *HNF1B* are driving both decreased fluid intelligence and renal function.

METHODS

Study subjects

Data from 450 993 participants of European ancestry from the UKB were analysed in this study. The UKB cohort is described in detail elsewhere.^{17 18} Phenotypes were derived using International Classification of Disease codes (ICD9 and ICD10) as well as from Hospital Episode Statistics (HES) data, serum and urinary biomarkers, and UKB-defined traits such as bipolar and major depression status or fluid intelligence. All phenotypes are detailed in online supplemental table 1. Participants withdrawn prior to the time of submission were excluded from analysis.

CNV calling

CNVs overlapping the 17q12 region were detected as outlined in Tuke *et al.*¹⁹ Briefly, SNP microarray data in the UKB were used to call CNVs using PennCNV V.1.0.4, with log R ratio (LRR) and B-allele frequency (BAF) values for 805 426 genome-wide probe sets provided by the UKB. All CNV calls were manually curated by inspecting plotted LRR and BAF. Large chromosomal aneuploidies and those with suspected mosaicism were excluded.

SNP association analysis

SNP genotypes were generated from the Affymetrix Axiom UKB array (~450 000 individuals) and the UK Biobank Lung Exome Variant Evaluation (UK BiLEVE) consortium array (~50 000 individuals) in 106 batches of ~4700 samples. This data set underwent extensive central quality control.¹⁸ High-quality imputed SNPs in the 17q12 region chr17:31827018–37956253 (GRCh37) (n=195 736) were extracted and then only those with minor allele frequency (MAF) >0.001 between the segmental microduplication regions (chr17:34442621–36711256) were included (n=535).

Exome sequencing

Variants detected using exome sequencing of 184 532 UKB participants were annotated using the Ensembl Variant Effect Predictor²⁰ with the LOFTEE plugin.²¹ Rare variants were included if they had a minor allele count (MAC) ≤30, LOFTEE high-confidence loss of function (LoF) or had a Combined

Annotation Dependent Depletion (CADD) score >30.²² The aligned sequence data for all variants meeting these criteria were visually inspected using the Integrative Genomics Viewer²³ (IGV) to remove likely false positives.

Phenotypes

Diabetes

Diabetes was defined as being one or more of the following: self-reported by participants or having an ICD9/ICD10 code for diabetes, or being on a diabetes treatment, or having glycated haemoglobin (HbA1c) ≥48 mmol/mol before recruitment.^{24 25}

Renal and liver disease

Estimated glomerular filtration rate (eGFR mL/min/1.73 m²) was calculated using the Chronic Kidney Disease Epidemiology Collaboration Creatinine-Cystatin Equation 2012 and was used to classify end-stage renal disease (ESRD), as well as broader chronic kidney disease categories: eGFR ≥60, eGFR <60 and eGFR <45. ESRD and eGFR <45 classifications included individuals who had renal replacement therapies such as kidney transplant or dialysis. Phenotypes for structural malformations of the kidney and ureter as well as any other structural malformation encoded in the HES data were also included in the analyses. A continuous measure of albumin to creatinine ratio (ACR) was derived from urinary microalbumin and creatinine. Individuals with an undetectable level of microalbumin were included as the minimum value in the UKB (6.7 mg/L), as has been done previously.²⁶ Serum biomarkers alkaline phosphatase (ALP), alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma glutamyl transferase (GGT), total bilirubin and direct bilirubin were used.

Fluid intelligence, general measures of functioning and neurodevelopmental/psychiatric disorders

Fluid intelligence, income, job class, educational qualification attainment level (qualification) and number of years in education were recorded. Fluid intelligence is measured as the number of questions (n=13) answered by participants in 2 min. Participants with diagnosed neurological or psychiatric disorders, such as intellectual disability or depression, were identified using HES codes. The developmental delay category combined neurodevelopmental disorders, bipolar, schizophrenia, depression, pervasive disorders, intellectual disability, epilepsy as well as structural malformations.

All phenotypes and associated codes are detailed in online supplemental table 1.

Association analysis

REGENIE (V.1.0.6.7)²⁷ was used for association testing and accounts for relatedness, among other factors. The null model for association testing was constructed using array genotypes with the following criteria: MAF ≥0.01, maximum Hardy-Weinberg Equilibrium p value of 1e⁻¹⁵, genotyping rate of 0.01 and missingness of 0.1 within individuals of European ancestry. We additionally pruned these SNPs based on the linkage structure within a white British subset of the UKB, with a maximum r² of 0.9, and then further only included variants with an MAC of 100. All continuous traits were single inverse normalised prior to regression testing, but all reported means were calculated from non-normalised data. Covariates included were age, sex and centre for all traits. The Strengthening the Reporting of Genetic Association Studies guidelines were used.²⁸

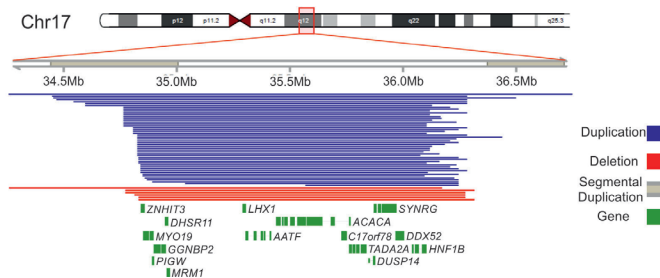


Figure 1 17q12 genomic region highlighting segmental microduplications, microdeletions and microduplications identified in this study. Only the unique detectable regions for microdeletions ($n=6$) and microduplications ($n=50$), based on CNV calling, are shown. The remaining identified microdeletions ($n=5$) and microduplications ($n=56$) share breakpoints, with one which is plotted.

Copy-number variants

17q12 microdeletions and microduplications were each coded as pseudo-heterozygous SNPs in pedigree format (PED) and the meta-files required for the association testing against the null model were generated using Plink (V.2.00a2LM).²⁹

17q12 single-nucleotide polymorphism

SNPs in the 17q12 region were tested against all phenotypes for associations. There were obvious areas where SNPs were absent from this region which correlated with the known segmental microduplications (figure 1). We applied a significance threshold of $p < 9 \times 10^{-5}$ (ie, $p = 0.05/535$) because we tested 535 SNPs in the region.

Exome variant burden testing

REGENIE-GENE²⁷ was used for burden testing of each of the 15 genes in the 17q12 region against all phenotypes. Burden tests were carried out for PTVs with and without pathogenic missense variants as well as single-gene burdens. No participants

with microdeletions also had a PTV in any of the other genes. We applied a significance threshold of $p < 3 \times 10^{-3}$ (ie, $p = 0.05/15$ genes tested).

RESULTS

Eleven 17q12 microdeletion and 106 17q12 microduplication carriers were identified in UKB

Using SNP array data, we found 11 microdeletions and 106 microduplications in the UKB cohort, ranging in size between 1.4–7.0 Mb and 1.2–7.0 Mb, respectively (figure 1). Ten microdeletions and 100 microduplications were identified in individuals of European ancestry and were included in downstream association analyses. Cohort clinical characteristics are included in online supplemental table 2.

Recurrent 17q12 microdeletions and microduplications are both associated with renal disease

eGFR ($\text{mL}/\text{min}/1.73 \text{ m}^2$) was consistently lower in individuals with microdeletions ($n=9$, mean=63.95; $\beta = -1.91$, 95% CI -2.45 to -1.37 , $p = 3 \times 10^{-12}$) and microduplications ($n=94$, mean=77.66; $\beta = -0.87$, 95% CI -1.04 to -0.71 , $p = 6 \times 10^{-25}$), relative to controls (mean=92.72) (table 1 and online supplemental table 2). Those with microduplications also had significantly higher eGFR than those with microdeletions ($\beta = 0.63$, 95% CI 0.47 to 0.79, $p < 7 \times 10^{-15}$).

We found 2913 out of 449 651 controls (0.65%) had moderately decreased renal function or ESRD (eGFR < 45 plus renal replacement therapies), which was a significantly lower proportion than individuals with either microdeletions (20%; OR=121.96, 95% CI 25.18 to 590.75, $p = 2 \times 10^{-5}$) or microduplications (4%; OR=6.25, 95% CI 2.33 to 16.75, $p = 3 \times 10^{-3}$) when compared with eGFR ≥ 45 as the controls. Furthermore, 1339 individuals out of 450 879 controls (0.3%) had ESRD in the UKB (age range 40.3–70.3). Again, this was a significantly lower proportion than observed in individuals with microdeletions ($n=1$; OR=63.29, 95% CI 8.710 to 459.92, $p < 0.003$,

Table 1 Genotype–phenotype association analysis for 17q12 microdeletions ($n=10$) and microduplications ($n=100$) in participants of European ancestry in the UK Biobank (continuous traits)

Phenotype	Deletions				Duplications			
	Beta	SE	95% CI	P value	Beta	SE	95% CI	P value
eGFR	-1.91	0.27	-2.18 to -1.64	3×10^{-12}	-0.87	0.08	-1.04 to -0.71	6×10^{-25}
ACR	0.45	0.30	-0.14 to 1.04	1×10^{-1}	0.40	0.10	0.21 to 0.59	4×10^{-5}
Urea	1.84	0.31	1.23 to 2.45	3×10^{-9}	0.68	0.10	0.49 to 0.87	1×10^{-12}
Total protein	-0.27	0.32	-0.90 to 0.35	4×10^{-1}	0.36	0.10	0.16 to 0.56	3×10^{-4}
HbA1c	1.26	0.28	0.71 to 1.81	7×10^{-6}	0.11	0.09	-0.07 to 0.29	3×10^{-1}
CRP	-0.47	0.31	-1.08 to 0.14	1×10^{-1}	0.44	0.10	0.25 to 0.63	4×10^{-6}
ALP	1.52	0.29	0.95 to 2.10	2×10^{-7}	0.07	0.09	-0.11 to 0.25	4×10^{-1}
AST	0.83	0.31	0.22 to 1.43	7×10^{-3}	0.21	0.10	0.03 to 0.40	3×10^{-2}
ALT	1.15	0.30	0.55 to 1.74	2×10^{-4}	0.03	0.09	-0.16 to 0.21	8×10^{-1}
GGT	0.81	0.29	0.24 to 1.38	5×10^{-3}	0.09	0.09	-0.09 to 0.26	3×10^{-1}
FI	-0.30	0.43	-1.15 to 0.54	5×10^{-1}	-0.59	0.16	-0.91 to -0.27	3×10^{-4}
Years in education	-0.85	0.30	-1.44 to -0.27	4×10^{-3}	-0.32	0.10	-0.51 to -0.13	8×10^{-4}
Job class	0.64	0.49	-0.32 to 1.61	2×10^{-1}	0.53	0.13	0.28 to 0.79	5×10^{-5}
Income	-1.15	0.35	-1.82 to -0.47	9×10^{-4}	-0.23	0.10	-0.44 to -0.03	2×10^{-2}
TDI	0.73	0.30	0.15 to 1.32	1×10^{-2}	0.24	0.10	0.05 to 0.43	9×10^{-3}
Qualifications	-0.86	0.30	-1.45 to -0.27	4×10^{-3}	-0.29	0.10	-0.48 to -0.11	2×10^{-3}

See online supplemental table 1 for phenotype details.

ACR, urinary albumin creatinine ratio; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; beta, regression coefficient; CRP, C reactive protein; eGFR, estimated glomerular filtration rate ($\text{mL}/\text{min}/1.73 \text{ m}^2$); Chronic Kidney Disease Epidemiology Collaboration Creatinine-Cystatin 2012; FI, fluid intelligence; GGT, gamma glutamyl transferase; HbA1c, glycated haemoglobin; TDI, Townsend Deprivation Index.

Table 2 Genotype–phenotype association analysis for 17q12 microdeletions (n=10) and microduplications (n=100) in participants of European ancestry in the UK Biobank (binary traits)

Phenotype	Deletions				Duplications			
	OR	SE	95% CI	P value	OR	SE	95% CI	P value
Diabetes	43.74	2.02	11.06 to 172.94	2×10^{-7}	1.37	1.49	0.63 to 2.98	4×10^{-1}
ESRD	63.29	2.75	8.71 to 459.92	3×10^{-3}	9.46	1.93	2.61 to 34.31	9×10^{-3}
eGFR <60	213.9	2.24	44.12 to 1036.96	2×10^{-9}	8.47	1.34	4.76 to 15.05	1×10^{-9}
eGFR <45	121.96	2.24	25.18 to 590.75	2×10^{-5}	6.24	1.65	2.33 to 16.75	3×10^{-3}
ID	155.94	2.69	22.37 to 1087.02	6×10^{-4}	0.37	17.23	1×10^{-3} to 96.92	7×10^{-1}
Psych	2.01	1.97	0.53 to 7.62	3×10^{-1}	1.75	1.23	1.17 to 2.61	7×10^{-3}
DD	1.49	1.91	0.42 to 5.32	5×10^{-1}	1.73	1.23	1.15 to 2.61	7×10^{-3}

See online supplemental table 1 for phenotype details.

DD, developmental delay; eGFR, estimated glomerular filtration rate (mL/min/1.73 m²; Chronic Kidney Disease Epidemiology Collaboration Creatinine-Cystatin 2012); ESRD, end-stage renal disease; GP, General Practitioner; HCP, Health Care Professional; ID, diagnosed intellectual disability; Psych, visit GP or HCP for psychiatric disorder.

age=43.25) or microduplications (n=2; OR=9.46, 95% CI 2.61 to 34.31, p=0.01, ages=65.5 and 68.1) (table 2 and figure 2).

Individuals with microdeletions had lower urinary ACR (mg/mmol) than those with microduplications ($\beta=-0.32$, 95% CI -0.50 to -0.14 , p= 5×10^{-4}) when compared directly. When compared with the background population (mean=1.98), ACR in microduplications was significantly raised (mean=5.41, $\beta=0.40$, 95% CI 0.21 to 0.59, p= 4×10^{-5}) but not in microdeletions (mean=2.32, $\beta=0.45$, 95% CI -0.14 to 1.03, p=0.13). Three participants in the microdeletion cohort had an ACR >60 mg/mmol, of whom one had ESRD and another had a moderately decreased eGFR of 51 mL/min/1.73 m².

Recurrent 17q12 microdeletions cause diabetes and abnormal liver function tests, but microduplications do not

As has been published previously, we found that 17q12 microdeletions were strongly associated with diabetes (OR=43.74, 95% CI 11.06 to 172.94, p= 2×10^{-7}), whereas the reciprocal microduplications were not (OR=1.37, 95% CI 0.63 to 2.98, p=0.4).

Similarly, microdeletions were associated with raised liver enzymes: ALP (U/L) (mean=196.9, $\beta=1.52$, 95% CI 0.95 to 2.1, p= 2×10^{-7}), ALT (U/L) (mean=49.24, $\beta=1.15$, 95% CI 0.55 to 1.74, p= 2×10^{-4}), AST (U/L) (mean=65.74, $\beta=0.83$, 95% CI 0.22 to 1.43, p= 7×10^{-3}) and GGT (U/L) (mean=112.0, $\beta=0.81$, 95% CI 0.24 to 1.38, p= 5×10^{-3}). Direct bilirubin and

total bilirubin were not significantly different in either microdeletions ($\beta=0.26$, 95% CI -0.43 to 0.94, p=0.46; $\beta=-0.17$, 95% CI -0.68 to 0.35, p=0.53) or microduplications ($\beta=-0.02$, 95% CI -0.20 to 0.16, p=0.84; $\beta=0.05$, 95% CI -0.11 to 0.21, p=0.56) (figure 2).

Cognitive ability is negatively affected in both 17q12 microdeletions and microduplications

Fluid intelligence was significantly lower in 17q12 recurrent microduplications ($\beta=-0.59$, 95% CI -0.91 to -0.27 , p= 3×10^{-4}) but not in the reciprocal microdeletions ($\beta=-0.30$, 95% CI -1.15 to 0.54, p=0.48), although the CIs overlapped. There was also an association with fewer years in education (microdeletions: $\beta=-0.85$, 95% CI -1.44 to -0.27 , p= 4×10^{-3} ; microduplications: $\beta=-0.32$, 95% CI -0.51 to -0.13 , p= 8×10^{-4}) as well as lower educational attainment (microdeletions: $\beta=-0.86$, 95% CI -1.45 to -0.27 , p= 4×10^{-3} ; microduplications: $\beta=-0.29$, 95% CI -0.48 to -0.11 , p= 2×10^{-3}). Additionally, microdeletions are associated with lower income ($\beta=-1.15$, 95% CI -1.82 to -0.47 , p= 9×10^{-4}) and microduplications associated with lower job class ($\beta=0.53$, 95% CI 0.28 to 0.79, p= 5×10^{-5}).

Duplications are associated with developmental delay (OR=1.73, 95% CI 1.15 to 2.61, p= 2×10^{-3}); however, microdeletions are not (OR=1.49, 95% CI 0.42 to 5.32, p=0.54), although the CIs overlapped. Microdeletions are more likely to result in a diagnosed intellectual disability (OR=155.94, 95% CI 22.37 to 1087.02, p= 6×10^{-4}). Individuals with a microduplication were also more likely to visit either a general practitioner or a psychiatrist for psychiatric disorder (OR=1.75, 95% CI 1.17 to 2.61, p= 2×10^{-3}) (figure 2).

SNP associations in the 17q12 region suggest that genes other than *HNF1B* may contribute to the observed phenotypes

We identified a genomic risk locus at *HNF1B* for diabetes (NC_000017.11:g.37741165C>G, rs7501939, p= 6×10^{-17}), increased ALT (NC_000017.11:g.37713312C>T, rs17138478, p= 6×10^{-25}), AST (NC_000017.11:g.37713312C>T, rs17138478, p= 4×10^{-12}) and GGT (NC_000017.11:g.37717101A>G, rs718961, p= 2×10^{-32}). For eGFR, two risk loci were identified: one around *HNF1B* (NC_000017.11:g.37741642T>G, rs12601991, p= 4×10^{-21}), as expected, and another at the other end of the deletion region (NC_000017.11:g.36558947T>C, rs12150665, p= 4×10^{-15}), encompassing five nearby genes (figure 3). Furthermore, this

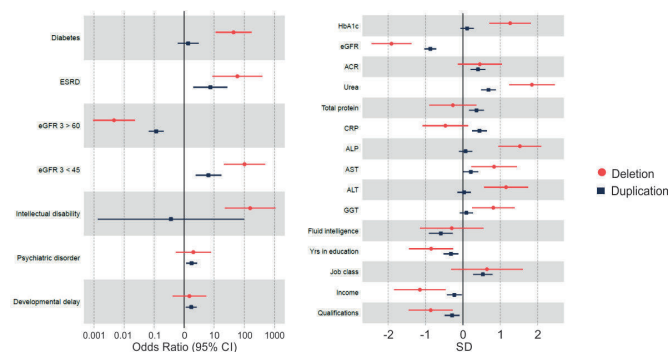


Figure 2 Forest plots of 17q12 microdeletions and microduplication disease associations in the UK Biobank. Left panel: binary traits; right panel: inverse normalised continuous traits. Note that SD below 0 represents negative associations. ACR, urinary albumin to creatinine ratio; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; CRP, C reactive protein; eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; GGT, gamma glutamyl transferase; HbA1c, glycated haemoglobin.

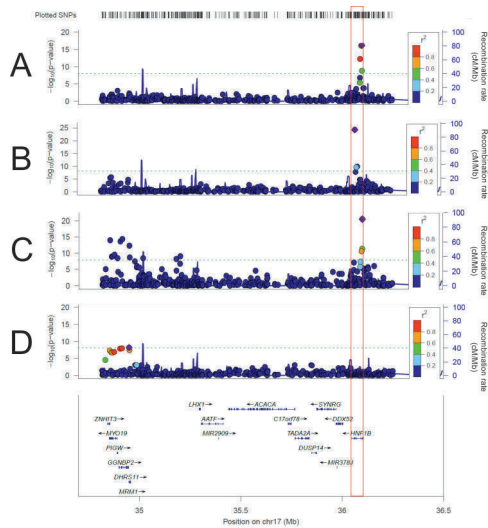


Figure 3 LocusZoom plots of GWAS SNP testing: (A) diabetes, (B) alanine aminotransferase, (C) estimated glomerular filtration rate and (D) fluid intelligence. The red vertical box indicates *HNF1B* location. The green dashed line indicates genome-wide significance at 1×10^{-8} .

second locus was also associated with decreased fluid intelligence (NC_000017.11:g.36593168T>G, rs1005552, $p=6 \times 10^{-9}$), while the *HNF1B* locus was not. This result suggests that, although *HNF1B* is the primary driver of decreased eGFR, it may not be the sole contributor and that a gene or genes in this second locus may explain the neurocognitive phenotypes (figure 3).

Colocalisation analysis between eGFR and fluid intelligence for all the SNPs in the region ($n=535$) predicted that both traits associate with the region but with different causal variants (posterior probability=0.99). When testing each genomic risk locus separately, the 5' locus ($n=199$ SNPs) predicted that both traits are associated and share a single causal variant (posterior probability=0.995), whereas the 3' locus ($n=336$) indicated that only eGFR was associated with the region (posterior probability=0.98).

Burden testing of 17q12 genes does not provide evidence of individual gene contributions to microdeletion phenotypes

Two PTVs and one pathogenic missense variant in *HNF1B* were identified in four UKB participants of European ancestry, one of whom had diabetes ($p=0.06$). These variants were associated with increased ACR ($\beta=1.66$, 95% CI 0.73 to 2.58, $p=4 \times 10^{-4}$; background: mean=17.53, range=1.32–18 932.28, median=9.80; *HNF1B* pathogenic background: mean=53.83, range=13.09–105.04, median=48.61). We also tested the associations with PTVs and rare missense variants if they had an MAC ≤ 30 , LOFTEE²¹ high-confidence LoF or a CADD score > 30 ²² in the 14 other genes in the 17q12 region. None of these variant groups was associated with decreased eGFR. *ZNFHIT3* (38 PTVs and 13 missense heterozygotes) was associated with structural renal abnormalities (OR=52.18, 95% CI 7.59 to 358.83, $p=4 \times 10^{-4}$); one individual had a horseshoe kidney (89 in UKB), one had a congenital bile duct malformation (18 in UKB) and one had an unspecified endocrine congenital malformation (121 in UKB). We also found associations between rare PTV/missense variants in *C17orf78* with elevated cystatin C ($\beta=0.51$, 95% CI 0.23 to 0.80, $p=3 \times 10^{-4}$) and *SYNRG* with

increased risk of bipolar disorder (OR=27.45, 95% CI 3.03 to 248.61, $p=8 \times 10^{-4}$).

DISCUSSION

We have shown that the prevalence of 17q12 recurrent microduplications is 1 in 4607 ($n=106$) and 1 in 44 398 for microdeletions ($n=11$) in a population-based cohort of 488 377 individuals in the UKB. We have confirmed that microdeletions associate with both diabetes and raised liver enzymes, whereas microduplications do not, consistent with previous publications in smaller cohorts.^{12 13 30–32}

Both microdeletions and microduplications are associated with decreased eGFR, despite only a few individuals reaching ESRD ($n=1$ and $n=2$, respectively). Microdeletions show a greater effect size than microduplications, which is consistent with other shared phenotypes.^{3 6} Heterozygous variants in *HNF1B* are one of the most common monogenic causes of developmental kidney disease.³³ Rare cases of renal malformations have been reported in microduplication carriers,³⁴ but systematic imaging of affected individuals has not been published so far, to the best of our knowledge. One of the limitations of this work was the small number of MRI results available for UKB participants, particularly for a condition like *HNF1B*-associated disease, where structural abnormalities are a key phenotypic feature. 17q12 microdeletions are known to cause diabetes and renal disease.¹ In this cohort we identify one individual with a microdeletion who has diabetes and ESRD, and five individuals with microdeletions who have diabetes and no ESRD. ESRD is a life-threatening disease status and represents severe progression of the observed 17q12 deletion syndrome. It is therefore not surprising that not all participants have late-stage disease progression, such as ESRD, particularly in a population cohort where participants would typically be assumed to be in 'good health' to be able to participate. We found an association with 17q12 microduplications and ESRD, but not diabetes. Diabetes can itself cause ESRD and we identify one individual with a microduplication who has diabetes and ESRD and one individual with a microduplication who has ESRD without diabetes. It does remain a possibility that the ESRD observed in the individuals with a microduplication could be caused by their diabetes. Given the low numbers of affected individuals, untangling the intricacies of these potential comorbidities and the underlying genetic drivers would be a suitable topic for further research.

Microdeletions and microduplications cause neurodevelopmental disorders. Microduplications are also associated with psychiatric disorders.¹⁴ We show intelligence traits are decreased in individuals with both microdeletions and microduplications, with microdeletions being more severe. Microdeletions were not associated with decreased fluid intelligence, but this likely reflects low statistical power. One individual with a microdeletion had mental retardation (ICD10: F70–79). This was one of only 144 participants in the UKB cohort reported to have this diagnosis. Microduplications were associated with developmental delay with a similar effect size to microdeletions. Microduplications were associated with visiting a healthcare professional for a psychiatric disorder, but not for individual psychiatric traits. This is likely due to the challenge of diagnosing neurodevelopmental and psychiatric conditions and the low incidence in the UKB due to recruitment bias. We found a genomic risk locus for fluid intelligence which was distant from *HNF1B*, reaffirming previous findings that other genes are likely responsible for neurodevelopmental phenotypes in 17q12 microdeletion and microduplication carriers.^{14 35}

Intragenic *HNF1B* pathogenic variants cause lower eGFR than 17q12 microdeletions, suggesting a dominant negative effect of the former.³⁶ Association testing of common SNPs in the region identified a second genomic risk locus for eGFR, with an intronic SNP in *GGNBP2* (NC_000017.11:g.36558947T>C, rs12150665) as the most significant. Gene burden testing associated *ZNHIT3* with an increased likelihood of structural abnormalities, but also *C17orf78* with increased cystatin C. Overexpression of *ACACA* causes podocyte cell death in vitro and may therefore cause the decreased eGFR in microduplications. This mechanism would also support the association with microduplications and raised ACR. *ACACA* knockdown does not exhibit a significant protective effect.³⁷ Association and colocalisation analyses strongly suggest that there are two independent genetic effects across the 17q12 variable-copy region and that *HNF1B* is unlikely the sole contributor to renal phenotypes in both microdeletions and microduplications.

To date, 17q12 CNVs have largely been studied in selective cohorts, which has led to ascertainment bias. The UKB is a valuable resource that can be leveraged to discover genotype-phenotype associations but is enriched in participants of higher socioeconomic stratification and lower rates of disease. This means studying rare genetic variation underlying complex, and often severe, disease presentation is more challenging. Specifically, the rarer the genetic variation and the more severe the phenotype, the less likely it is to be observed in a population cohort. Similarly, some phenotypes are more easily captured and documented in a population cohort than others. Examples in this study include psychiatric conditions but also the lack of MRI in affected individuals. The latter is a drawback in a condition like *HNF1B*-associated disease, where structural abnormalities are one of the main phenotypic features. It can also be challenging to identify the age of onset of disorders and whether they occur in parallel or isolation. Despite this, carriers of 17q12 microdeletions and microduplications are of a similar age at recruitment to the general population (online supplemental table 2).

CONCLUSIONS

We show that both 17q12 microdeletions and microduplications are associated with renal disease and provide evidence that *HNF1B* is unlikely to be the sole contributor to all associated phenotypes. This work highlights the utility of population cohorts with high-resolution genomic and phenotypic data, such as UKB, to study multigene disorders. As population data sets increase in size and the accuracy of variant detection and interpretation improves, we will be better able to characterise the genotype-phenotype associations in complex multisystem disorders like 17q12 deletion and microduplications.

Twitter Mark Gilchrist @gilchristN03

Acknowledgements The authors would like to acknowledge the use of the University of Exeter High-Performance Computing (HPC) facility in carrying out this work. The study was conducted using the UK Biobank resource under application numbers 9072 and 49847.

Contributors Conceptualisation: SC, MNW, CW, KP, ATH, RAO. Data curation: SC, MNW, KP, MT. Formal analysis: SC, MT. Funding acquisition: MNW. Investigation: SC, MT, KP. Methodology: SC, MT, GH, RNB, ARW, RAO, KS. Project administration: SC, MNW, KP. Software: SC, GH, RNB, ARW. Supervision: RC, RAO. Visualisation: SC. Writing—original draft: SC. Writing—review and editing: SC, MNW, CW, KP, RNB, GH, MG, RAO. Guarantor: MW.

Funding This work was supported by Diabetes UK (no: 19/0005994) and MRC (grant no: MR/T00200X/1). KP is funded by the Wellcome Trust (no: 219606/Z/19/Z). ATH is supported by a Wellcome Trust Senior Investigator Award (no: WT098395/Z/12/Z). RAO is funded by a Diabetes UK Harry Keen Fellowship (16/0005529).

Disclaimer The views expressed are those of the authors and are not necessarily those of the Wellcome Trust, Department of Health, NHS or NIHR.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study involves human participants and was approved by the North West Centre for Research Ethics Committee (11/NW/0382). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

ORCID iDs

Stuart Cannon <http://orcid.org/0000-0002-5108-7615>

Mark Gilchrist <http://orcid.org/0000-0002-0284-6048>

REFERENCES

- Rasmussen M, Vestergaard EM, Graakjaer J, Petkov Y, Bache I, Fagerberg C, Kibaek M, Svaneby D, Petersen OB, Brasch-Andersen C, Sunde L. 17Q12 deletion and duplication syndrome in Denmark-A clinical cohort of 38 patients and review of the literature. *Am J Med Genet A* 2016;170:2934–42.
- Izzi C, Dordoni C, Econimo L, Delbarba E, Grati FR, Martin E, Mazza C, Savoldi G, Rampoldi L, Alberici F, Scolari F. Variable expressivity of HNF1B nephropathy, from renal cysts and diabetes to medullary sponge kidney through tubulo-interstitial kidney disease. *Kidney Int Rep* 2020;5:2341–50.
- Clissold RL, Hamilton AJ, Hattersley AT, Ellard S, Bingham C. HNF1B-associated renal and extra-renal disease—an expanding clinical spectrum. *Nat Rev Nephrol* 2015;11:102–12.
- Kamath A, Linden SC, Evans FM, Hall J, Jose SF, Spillane SA, Hardie ADR, Morgan SM, Pilz DT. Chromosome 17q12 duplications: further delineation of the range of psychiatric and clinical phenotypes. *Am J Med Genet B Neuropsychiatr Genet* 2018;177:520–8.
- Ahn H-S, Kim JH, Jeong H, Yu J, Yeom J, Song SH, Kim SS, Kim U, Kim K. Differential urinary proteome analysis for predicting prognosis in type 2 diabetes patients with and without renal dysfunction. *Int J Mol Sci* 2020;21:4236.
- Dubois-Laforgue D, Cornu E, Saint-Martin C, Coste J, Bellanné-Chantelot C, Timsit J, Monogenic Diabetes Study Group of the Société Francophone du Diabète. Diabetes, Associated Clinical Spectrum, Long-term Prognosis, and Genotype/Phenotype Correlations in 201 Adult Patients With Hepatocyte Nuclear Factor 1B (*HNF1B*) Molecular Defects. *Diabetes Care* 2017;40:1436–43.
- Quintero-Rivera F, Woo JS, Bomberg EM, Wallace WD, Peredo J, Dipple KM. Duodenal atresia in 17q12 microdeletion including HNF1B: a new associated malformation in this syndrome. *Am J Med Genet A* 2014;164A:3076–82.
- Kettunen JLT, Parviainen H, Miettinen PJ, Färkkilä M, Tamminen M, Salonen P, Lantto E, Tuomi T. Biliary anomalies in patients with HNF1B diabetes. *J Clin Endocrinol Metab* 2017;102:2075–82.
- Kotalova R, Dusatkova P, Cinek O, Dusatkova L, Dedic T, Seeman T, Lebl J, Pruhova S. Hepatic phenotypes of HNF1B gene mutations: a case of neonatal cholestasis requiring portoenterostomy and literature review. *World J Gastroenterol* 2015;21:2550–7.
- Edghill EL, Oram RA, Owens M, Stals KL, Harries LW, Hattersley AT, Ellard S, Bingham C. Hepatocyte nuclear factor-1beta gene deletions—a common cause of renal disease. *Nephrol Dial Transplant* 2008;23:627–35.
- Bellanné-Chantelot C, Clauin S, Chauveau D, Collin P, Daumont M, Douillard C, Dubois-Laforgue D, Dusselier L, Gautier J-F, Jadoul M, Laloi-Michel M, Jacquesson L, Larger E, Louis J, Nicolino M, Subra J-F, Wilhelm J-M, Young J, Velho G, Timsit J. Large genomic rearrangements in the hepatocyte nuclear factor-1beta (TCF2) gene are the most frequent cause of maturity-onset diabetes of the young type 5. *Diabetes* 2005;54:3126–32.

- 12 Omura Y, Yagi K, Honoki H, Iwata M, Enkaku A, Takikawa A, Kuwano T, Watanabe Y, Nishimura A, Liu J, Chujo D, Fujisaka S, Enya M, Horikawa Y, Tobe K. Clinical manifestations of a sporadic maturity-onset diabetes of the young (MODY) 5 with a whole deletion of HNF1B based on 17q12 microdeletion. *Endocr J* 2019;66:1113–6.
- 13 Bingham C, Bulman MP, Ellard S, Allen LI, Lipkin GW, Hoff WG, Woolf AS, Rizzoni G, Novelli G, Nicholls AJ, Hattersley AT. Mutations in the hepatocyte nuclear factor-1beta gene are associated with familial hypoplastic glomerulocystic kidney disease. *Am J Hum Genet* 2001;68:219–24.
- 14 Clissold RL, Shaw-Smith C, Turnpenny P, Bunce B, Bockenbauer D, Kerecuk L, Waller S, Bowman P, Ford T, Ellard S, Hattersley AT, Bingham C. Chromosome 17q12 microdeletions but not intragenic HNF1B mutations link developmental kidney disease and psychiatric disorder. *Kidney Int* 2016;90:203–11.
- 15 Luigetti M, Del Grande A, Conte A, Lo Monaco M, Bisogni G, Romano A, Zollino M, Rossini PM, Sabatelli M. Clinical, neurophysiological and pathological findings of HNPP patients with 17p12 deletion: a single-centre experience. *J Neurol Sci* 2014;341:46–50.
- 16 Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, Moreno-De-Luca D, Moreno-De-Luca A, Mülle JG, Warren ST, Richard G, Compton JG, Fuller AE, Gliem TJ, Huang S, Collinson MN, Beal SJ, Ackley T, Pickering DL, Golden DM, Aston E, Whitby H, Shetty S, Rossi MR, Rudd MK, South ST, Brothman AR, Sanger WG, Iyer RK, Crolla JA, Thorland EC, Aradhya S, Ledbetter DH, Martin CL. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. *Genet Med* 2011;13:777–84.
- 17 Ruth KS, Day FR, Hussain J, Martínez-Marchal A, Aiken CE, Azad A, Thompson DJ, Knoblochova L, Abe H, Tarry-Adkins JL, Gonzalez JM, Fontanillas P, Claringbould A, Bakker OB, Sulem P, Walters RG, Terao C, Turon S, Horikoshi M, Lin K, Onland-Moret NC, Sankar A, Hertz EPT, Timshel PN, Shukla V, Borup R, Olsen KW, Aguilera F, Ferrer-Roda M, Huang Y, Stankovic S, PRHJ T, Ahearn TU, Alizadeh BZ, Naderi E, Andrusis IL, Arnold AM, Aronson KJ, Augustinsson A, Bandinelli S, Barbieri CM, Beaumont RN, Becher H, Beckmann MW, Benonisdottir S, Bergmann S, Bochud M, Boerwinkle E, Bojesen SE, Bolla MK, Boomsma DI, Bowker N, Brody JA, Broer L, Buring JE, Campbell A, Campbell H, Castela JE, Catamo E, Chanock SJ, Chenevix-Trench G, Ciullo M, Corre T, Couch FJ, Cox A, Crisponi L, Cross SS, Cucca F, Czene K, Smith GD, EJCN deG, de Mutsert R, Vivo D I, Demerath EW, Dennis J, Dunning AM, Dwek M, Eriksson M, Esko T, Fasching PA, Faul JD, Ferrucci L, Franceschini N, Frayling TM, Gago-Dominguez M, Mezzavilla M, Garcia-Closas M, Gieger C, Giles GG, Grallert H, Gudbjartsson DF, Gudnason V, Guénel P, Haiman CA, Håkansson N, Hall P, Hayward C, He C, He W, Heiss G, Hoffding MK, Hopper JL, Hottenga JJ, Hu F, Hunter D, Ikram MA, Jackson RD, Joaquin MDR, John EM, Joshi PK, Karasik D, Kardias LR, Kartsonaki C, Karlsson R, Kitahara CM, Kolcic I, Kooperberg C, Kraft P, Kurian AW, Kutalik Z, La Bianca M, LaChance G, Langenberg C, Launer LJ, Laven JSE, Lawlor DA, Le Marchand L, Li J, Lindblom A, Lindstrom S, Lindstrom T, Linet M, Liu Y, Liu S, Luan J, Mägi R, Magnusson PKE, Mangino M, Mannermaa A, Marco B, Marten J, Martin NG, Mbarek H, McKnight B, Medland SE, Meisinger C, Meitinger T, Menni C, Metspalu A, Milani L, Milne RL, Montgomery GW, Mook-Kanamori DO, Mulas A, Mulligan AM, Murray A, Nalls MA, Newman A, Noordam R, Nutile T, Nyholt DR, Olshan AF, Olsson H, Painter JN, Patel AV, Pedersen NL, Perjakova N, Peters A, Peters U, Pharoah PDP, Polasek O, Porcu E, Psaty BM, Rahman I, Rennert G, Rennert HS, Ridker PM, Ring SM, Robino A, Rose LM, Rosendaal FR, Rossouw J, Rudan I, Ruedi R, Ruggiero D, Sala CF, Saloustros E, Sandler DP, Sanna S, Sawyer EJ, Sarnowski C, Schlessinger D, Schmidt MK, Schoemaker MJ, Schraut KE, Scott C, Shekari S, Shrikhande A, Smith AV, Smith BH, Smith JA, Sorice R, Southey MC, Spector TD, Spinelli JJ, Stampfer M, Stöckl D, van Meurs JBJ, Strauch K, Styrkarsdottir U, Sverdlow AJ, Tanaka T, Teras LR, Teumer A, Borsteinsdottir U, Timpon NJ, Toniolo D, Traglia M, Troester MA, Truong T, Tyrrell J, Uitterlinden AG, Ulivi S, Vachon CM, Vitart V, Völker U, Vollenweider P, Völzke H, Wang Q, Wareham NJ, Weinberg CR, Weir DR, Wilcox AN, van Dijk KW, Willemssen G, Wilson JF, Wolfenbittel BHR, Wolk A, Wood AR, Zhao W, Zygumt M. Genetic insights into biological mechanisms governing human ovarian ageing. *Nature* 2021;596:393–7.
- 18 Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, Cortes A, Welsh S, Young A, Effingham M, McVean G, Leslie S, Allen N, Donnelly P, Marchini J. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562:203–9.
- 19 Tuke M, Tyrrell J, Ruth KS, Beaumont RN, Wood AR, Murray A, Frayling TM, Weedon MN, Wright CF. Large copy-number variants in UK Biobank caused by clonal hematopoiesis may confound penetrance estimates. *Am J Hum Genet* 2020;107:325–9.
- 20 McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. The Ensembl variant effect predictor. *Genome Biol* 2016;17:122.
- 21 Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Pober T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferreria S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Neale BM, Daly MJ, MacArthur DG, Genome Aggregation Database Consortium. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020;581:434–43.
- 22 Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* 2014;46:310–5.
- 23 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–6.
- 24 American Diabetes Association. 2. Classification and Diagnosis of Diabetes: *Standards of Medical Care in Diabetes-2020*. *Diabetes Care* 2020;43:S14–31.
- 25 World Health Organisation. *Use of glycated haemoglobin (HbA1c) in the diagnosis of diabetes mellitus: abbreviated report of a who consultation*. Geneva: World Health Organization, 2011.
- 26 Casanova F, Wood AR, Yaghoobkar H, Beaumont RN, Jones SE, Gooding KM, Aizawa K, Strain WD, Hattersley AT, Khan F, Shore AC, Frayling TM, Tyrrell J. A Mendelian randomization study provides evidence that adiposity and dyslipidemia lead to lower urinary Albumin-to-Creatinine ratio, a marker of microvascular function. *Diabetes* 2020;69:1072–82.
- 27 Mbatchou J, Barnard L, Backman J, Marcketta A, Kosmicki JA, Ziyatdinov A, Benner C, O'Dushlaine C, Barber M, Boutkov B, Habegger L, Ferreira M, Baras A, Reid J, Abecasis G, Maxwell E, Marchini J. Computationally efficient whole-genome regression for quantitative and binary traits. *Nat Genet* 2021;53:1097–103.
- 28 Little J, Higgins JPT, Ioannidis JPA, Moher D, Gagnon F, von Elm E, Khoury MJ, Cohen B, Davey-Smith G, Grimshaw J, Scheet P, Gwinn M, Williamson RE, Zou GY, Hutchings K, Johnson CY, Tait V, Wiens M, Golding J, van Duijn C, McLaughlin J, Paterson A, Wells G, Fortier I, Freedman P, Zecevic M, King R, Infante-Rivard C, Stewart A, Birkett N. Strengthening the Reporting of Genetic Association Studies. Strengthening the reporting of genetic association studies (STREGA): an extension of the STROBE statement. *PLoS Med* 2009;6:e22.
- 29 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
- 30 Frayling TM, Bulman MP, Ellard S, Appleton M, Dronsfield MJ, Mackie AD, Baird JD, Kaisaki PJ, Yamagata K, Bell GI, Bain SC, Hattersley AT. Mutations in the hepatocyte nuclear factor-1alpha gene are a common cause of maturity-onset diabetes of the young in the U.K. *Diabetes* 1997;46:720–5.
- 31 Mefford HC, Clauin S, Sharp AJ, Moller RS, Ullmann R, Kapur R, Pinkel D, Cooper GM, Ventura M, Ropers HH, Tommerup N, Eichler EE, Bellanne-Chantelot C. Recurrent reciprocal genomic rearrangements of 17q12 are associated with renal disease, diabetes, and epilepsy. *Am J Hum Genet* 2007;81:1057–69.
- 32 Tattersall RB. Mild familial diabetes with dominant inheritance. *Q J Med* 1974;43:339–57.
- 33 Thomas R, Sanna-Cherchi S, Warady BA, Furth SL, Kaskel FJ, Gharavi AG. Hnf1B and Pax2 mutations are a common cause of renal hypodysplasia in the CKiD cohort. *Pediatr Nephrol* 2011;26:897–903.
- 34 Faguer S, Chassaing N, Bandin F, Prouheze C, Arveiler B, Rooryck C, Nogier M-B, Chauveau D, Calvas P, Decramer S. A 17q12 chromosomal duplication associated with renal disease and esophageal atresia. *Eur J Med Genet* 2011;54:e437–40.
- 35 Vasileiou G, Hoyer J, Thiel CT, Schaefer J, Zapke M, Krumbiegel M, Kraus C, Zweier M, Uebe S, Ekici AB, Schneider M, Wiesener M, Rauch A, Faschingbauer F, Reis A, Zweier C, Popp B. Prenatal diagnosis of HNF1B-associated renal cysts: is there a need to differentiate intragenic variants from 17q12 microdeletion syndrome? *Prenat Diagn* 2019;39:1136–47.
- 36 Clissold RL, Harries LW, Ellard S, Bingham C, Hattersley AT. Comment on Dubois-Laforgue et al. Diabetes, Associated Clinical Spectrum, Long-term Prognosis, and Genotype/Phenotype Correlations in 201 Adult Patients With Hepatocyte Nuclear Factor 1B (HNF1B) Molecular Defects. *Diabetes Care* 2017;40:1436–1443. *Diabetes Care* 2018;41:e7.
- 37 Kampe K, Sieber J, Orellana JM, Mundel P, Jehle AW. Susceptibility of podocytes to palmitic acid is regulated by fatty acid oxidation and inversely depends on acetyl-CoA carboxylases 1 and 2. *Am J Physiol Renal Physiol* 2014;306:F401–9.

Phenotype	ICD10/UKB code	Details
Renal disease (ESRD or eGFR <15)	N180	End stage renal disease
	N185	CKD stage 5
	T861	Complications of kidney transplant
	Z940	Kidney transplant status
	Z992	Dependence on renal dialysis
	Z492	Other dialysis
	Z491	Extracorporeal dialysis
	Z490	Preparatory care for dialysis
	T861	Complications of kidney transplant
Other kidney malformations	N27	Small kidney
	Q63	Other congenital malformations of kidney
Malformations of ureter	Q621	Congenital occlusion of ureter
	Q622	Congenital mega ureter
	Q623	Other obstructive defects of renal pelvis and ureter
	Q624	Agenesis of ureter
	Q625	Duplication of ureter
	Q626	Malposition of ureter
	Q627	Congenital vesico-uretero-renal reflux
	Q628	Other congenital malformations of ureter
Kidney Cysts	N281	Cyst of kidney, acquired
	Q612	Polycystic kidney, adult type
Structural malformations of the kidney	Q60	Renal agenesis and other reduction defects of kidney
	Q61	Cystic kidney disease
	Q620	Congenital hydronephrosis
	Q631	Lobulated, fused and horseshoe kidney
Family history of renal disease	Z841	Family history of disorders of kidney and ureter
Disorders of kidney or ureter	N288	Other specified disorders of kidney and ureter
	N289	Disorder of kidney and ureter, unspecified

eGFR	30700	Creatinine
	30720	Cystatin C
ACR	30500	Microalbumin in urine
	30510	Creatinine (enzymatic) in urine
Diabetes	30750	Glycated haemoglobin (HbA1c)
	20003	Medications - insulin, sulfonylureas, metformin
Years in education	22501	Years in education
Fluid intelligence	20016	Fluid intelligence
Income	26411	Income
Job class	20277	Job class
Qualifications	6138	Qualifications
Urinary biomarkers	30500	Microalbumin in urine
	30510	Creatinine (enzymatic) in urine
	30520	Potassium in urine
	30530	Sodium in urine
Serum biomarkers	30600	Albumin
	30610	Alkaline phosphatase
	30620	Alanine aminotransferase
	30630	Apolipoprotein A
	30640	Apolipoprotein B
	30650	Aspartate aminotransferase
	30660	Direct bilirubin
	30670	Urea
	30680	Calcium
	30690	Cholesterol
	30700	Creatinine
	30710	C-reactive protein
	30720	Cystatin C
30730	Gamma glutamyltransferase	

	30740	Glucose
	30750	Glycated haemoglobin (HbA1c)
	30760	HDL cholesterol
	30770	IGF-1
	30780	LDL direct
	30790	Lipoprotein A
	30800	Oestradiol
	30810	Phosphate
	30820	Rheumatoid factor
	30830	SHBG
	30840	Total bilirubin
	30850	Testosterone
	30860	Total protein
	30870	Triglycerides
	30880	Urate
	30890	Vitamin D
<hr/>		
Poor mental state	2030	Guilty feelings
	1950	Sensitivity/ hurt feelins
	1940	irritability
	2020	Loneliness, isolation
	1930	Miserableness
	1920	Mood swings
	1970	Nervous feelings
	2040	Risk taking
	2010	Suffer from 'nerves'
	1990	Tense/'highly strung'
	1980	Worrier/anxious feelings
	2000	Worry too long after embarrassment
<hr/>		
Intellectual disability	F70	Mild intellectual disabilities
	F71	Moderate intellectual disabilities

	F72	Severe intellectual disabilities
	F73	Profound intellectual disabilities
	F78	Other intellectual disabilities
	F79	Unspecified intellectual disabilities
	F800	Phonological disorder
	F801	Expressive language disorder
	F802	Mixed receptive-expressive language disorder
	F804	Speech and language developmental delay due to hearing loss
	F808	Other developmental disorders of speech and language
	F809	Developmental disorders of speech and language, unspecified
	F810	Specific reading disorder
	F812	Mathematics disorder
	F818	Other developmental disorders of scholastic skills
	F819	Developmental disorders of scholastic skills, unspecified
	F82	Specific developmental disorder of motor function
<hr/>		
Visited healthcare professional for psychiatric disorder	2090	Seen GP for nerves, anxiety, tension or depression
	2100	Seen a psychiatrist for nerves, anxiety, tension or depression
<hr/>		
Epilepsy	131048	Date G40 first reported (epilepsy)
	G400	Localisation related idiopathic epilepsy and epileptic syndromes with seizures of localised onset
	G401	Localisation related idiopathic epilepsy and epileptic syndromes with simple partial seizures
	G402	Localisation related idiopathic epilepsy and epileptic syndromes with complex partial seizures
	G403	Generalised idiopathic epilepsy and epileptic syndromes
	G404	Other generalised epilepsy and epileptic syndromes
	G408	Other epilepsy and recurrent seizures
	G409	Epilepsy, unspecified
	G40A	Absence epileptic syndrome

Schizophrenia	F20	Schizophrenia
	F21	schizotypal disorder
	F22	Delusional disorders
	F23	Brief psychotic disorder
	F24	Shared psychotic disorder
	F25	Schizoaffective disorders
	F28	Other psychotic disorder not due to a substance or known physiological condition
	F29	Unspecified psychosis not due to a substance or known physiological condition
Bipolar	F30	Manic episode
	F31	Bipolar disorder
	F32	Major depressive disorder, single episode
	F33	Major depressive disorder, recurrent
	F34	Persistent mood [affective] disorders
	F38	Other mood affective disorder
	F39	Unspecified mood affective disorder
	20122	Bipolar disorder status
Congenital malformation	Q0	Nervous system
	Q1	Eye, ear, face, neck
	Q2	Circulatory system
	Q3	Respiratory system, cleft lip and palate
	Q4	Digestive system
	Q5	Genital organs
	Q6	Urinary system
	Q7	Musculoskeletal system
	Q8	Other
	Q9	Chromosomal abnormalities, not elsewhere classified
Pervasive	F840	Autistic disorder
	F842	Rett's syndrome
	F843	Other childhood disintegrative disorder

	F845	Asperger's syndrome
	F848	Other pervasive developmental disorders
	F849	Pervasive developmental disorder, unspecified
Developmental delay	As above	Bipolar
	As above	Schizophrenia
	As above	Pervasive
	As above	Visited healthcare professional for psychiatric disorder
	As above	Intellectual disability
	As above	Malformation
	As above	Epilepsy

Supplementary table 1 – Phenotype definitions

Category	UKB background					17q12 Microdeletion					17q12 Microduplication				
	Mean	95%CI	n	%	Total	Mean	95%CI	n	%	Total	Mean	95%CI	n	%	Total
Participants	-	-	450,879	89.75	502,371	-	-	10	90.91	11	-	-	100	94.34	106
Sex (male)	-	-	206,140	45.72	450,879	-	-	5	50	10	-	-	6	60	100
Age (years)	57.28	57.26, 57.31	-	-	450,879	52.58	46.22-58.93	-	-	10	56.84	55.19-58.48	-	-	100
BMI	27.40	27.38-27.41	-	-	449,140	24.87	22.86-26.88	-	-	10	28.35	27.27-29.43	-	-	100
Diabetes	-	-	30,191	6.70	450,879	-	-	6	60	10	-	-	9	9	100
eGFR	92.72	92.67-92.77	-	-	429,395	63.95	50.59-77.32	-	-	9	77.66	73.96-81.37	-	-	94
ACR	1.98	1.95-2.01	-	-	437,935	2.32	0.62-4.03	-	-	10	5.41	1.67-9.15	-	-	95
ALP	83.59	83.52-83.67	-	-	429,948	196.9	33.63-360.24	-	-	9	85.13	80.06-90.21	-	-	94
AST	26.21	26.18-26.24	-	-	428,348	65.74	-12.62-144.11	-	-	9	27.50	26.06-28.93	-	-	94
ALT	23.54	23.94-23.58	-	-	429,777	49.24	12.42-86.07	-	-	9	24.79	22.25-27.33	-	-	94
GGT	37.37	37.24-37.50	-	-	429,706	112.0	10.33-213.72	-	-	9	40.82	34.78-46.86	-	-	94
Smoking	-	-	202,204	45.46	444,793	-	-	4	40	10	-	-	50	51.02	98
Obesity	-	-	110,326	24.47	450,879	-	-	0	0	10	-	-	30	30	100
CKD Family history	-	-	61	0.01	450,879	-	-	0	0	10	-	-	0	0	100
Hypertension	-	-	241,572	53.92	448,000	-	-	3	30	10	-	-	59	59	100
BP medication	-	-	92,981	20.77	447,703	-	-	3	30	10	-	-	29	29	100
Systolic BP	144.24	144.17-144.31	-	-	449,857	136.6	118.94-154.26	-	-	10	146.36	141.24-151.48	-	-	100
Diastolic BP	86.35	86.64-86.69	-	-	449,857	88.25	77.83-98.37	-	-	10	87.52	84.69-90.35	-	-	100
ESRD	-	-	1,339	0.30	450,879	-	-	1	10	10	-	-	2	2	100

Supplementary Table 2 – Cohort clinical characteristics. UKB background = EU ancestry excluding; 17q12 microdeletions, microduplications and *HNFB1B*

pathogenic mutations. Age = age at recruitment to UKB study. Smoking = if participants had ever smoked. Obesity calculated from BMI. Hypertension based on blood pressure medications and blood pressure. ESRD = eGFR < 15 plus participants with renal replacement therapies; ALP = alkaline phosphatase; AST = aspartate amino transferase; ALT = alanine aminotransferase; GGT = gamma-glutamyl transferase;