Exome-Wide Collapsing Analysis Reveals Novel Genes for Congenital Obstructive Uropathy

Juntao Ke[1], Dina F. Ahram[1], Yask Gupta[1], Tze Y. Lim[1], Gundula Povysil[2], Gina Jin[1], Ali G. Gharavi[1], Cathy Mendelsohon[2], David B. Goldstein[3], Simone Sanna-Cherchi[1]

[1]Division of Nephrology, Department of Medicine, Columbia University, New York, New York.
[3]Institute for Genomic Medicine, Columbia University, New York, New York.

Background

- As a subset of Congenital Anomalies of the Kidney and Urinary Tract (CAKUT), congenital obstructive uropathy (COU) is the most frequent abnormality of the urinary tract occurring in up to 2% of pregnancies, and constituting a leading cause of pediatric chronic kidney disease[1]. A number of developmental malformations constitute COU defects including ureteropelvic junction obstruction (UPJO), megaloureter / ureterovesical junction obstruction (UVJO), and congenital hydronephrosis not otherwise characterized. As such, COU represents a heterogeneous group of pathologies with variable overlap between each other, making the diagnosis and treatment often challenging.
- The gene mapping of COU has long been impeded by its clinical and genetic heterogeneity, its incomplete penetrance, and a genetic architecture likely far more complex than previously anticipated. Recent advances in high-throughput genomic technologies provide an opportunity to obtain fundamental insight for COU, allowing us to reclassify the subtypes based on genetic pathogenesis and leading to more precise diagnosis, therapy, and counseling.

Cohort Selection & Methods

- We conducted a exome sequencing (ES) study on 822 COU cases, encompassing three main classes of congenital urinary obstructions: a) UPJO (N=338), UVJO (UVJO / megaloureter; N=217), and COU not otherwise specified (COU-NOS; N=267) (**Fig. 1**).

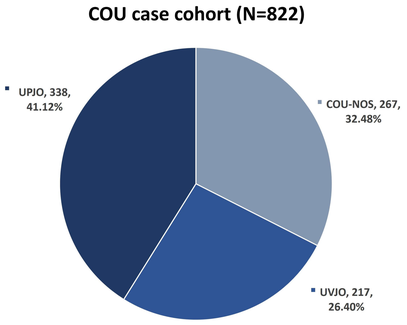

Fig. 1 We included 822 unrelated cases with clinically heterogeneous COU subphenotypes.. The primary COU subphenotypes are abbreviated as follow: UPJO: Ureteropelvic Junction Obstruction (n=338), UVJO: Ureterovesical Junction Obstruction / megaloureter (n=217), COU-NOS: COU not otherwise specified (n=267).

- To investigate the excess burden of rare genetic coding variants on COU, we performed an exome-wide collapsing analysis comparing the above 822 clinically ascertained cases and 23,958 matched controls with ES data, using our bioinformatics infrastructure Analysis Tool for Annotated Variants (ATAV, https://github.com/igm-team/atav)[2] at the Institute of Genomic Medicine (IGM). We used an original analytical pipeline for quality control and data harmonization in order to mitigate bias from population structure and sequencing platform heterogeneity.

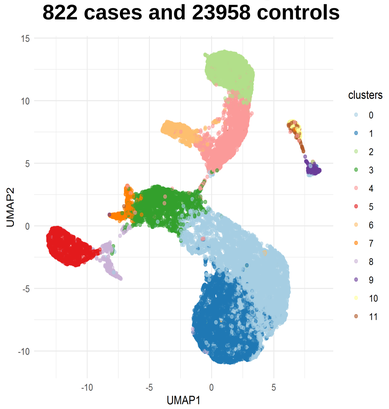

Fig. 2 The UMap clustering of our COU cohort.

- We first conducted coverage harmonization across genes or regions with different coverage across cases and controls. Principal Component Analysis (PCA) for dimensionality reduction was applied to capture population structure. Using the first 6 principal components (PCs) as input, we utilized the clustering method Louvain of community detection[3] for identifying clusters within the data that reflect the ethnicity of the samples. Our COU cohort was divided into 11 different clusters (**Fig. 2**). For each cluster with sufficient numbers of cases (N≥10) and controls, we performed coverage harmonization and gene-based collapsing to find genes with an enrichment of rare qualifying variants (QVs) in cases or controls. From the collapsing results of the individual clusters, we extracted the number of cases/controls with and without a QV per gene and use the Cochran-Mantel-Haenszel (CMH) test[4] for generating a global p-value.
- Analyses were conducted on the entire dataset and then for each of the three COU sub-phenotypes, before and after removal of cases harboring diagnostic/pathogenic Mendelian mutations and structural variants based on the American College of Medical Genetics and Genomics (ACMG) guidelines[5,6] for clinical variant interpretation[7].

Results

- In the analysis on the entire cohort of 822 COU cases and 23,958 controls, the top signal was for *ANG*, encoding angiogenin, in the dominant protein-truncating-variant (PTV) model (P= 3.25 × 10^{-5}; OR= 24.40). The signal improved after removal of solved cases, approaching exome-wide significance (P= 8.28 × 10^{-6}; OR= 36.37: 667 cases and 13258 controls), and supporting candidacy for this gene.
- From the COU subtype analysis, we found the following suggestive signals: *DMTF1* in the dominant ultrarare model (P= 1.97 × 10^{-4}; OR=infinite) and *BLVRB* in the recessive autosomal model (P= 1.97 × 10^{-4}; OR=infinite) for COU-NOS; *KLRD1* in the dominant PTV model (P= 9.32 × 10^{-5}; OR=infinite) and *EXOSC2* in the dominant ultrarare model (P= 1.86 × 10^{-5}; OR=infinite) for UPJO; and *MMP15* in the dominant rare model (P= 6.86 × 10^{-5}; OR=16.29) for UVJO.

Gene	Model	OR (95%CI)	*P*-value	Qualified Cases	Qualified Controls	Cohort (Case/Control)
ANG	Dominant PTV	24.40 (5.30-105.74)	3.25 × 10^{-5}	5	6	Total (774/16370)
ANG	Dominant PTV	36.37 (7.29-193.43)	8.28 × 10^{-6}	5	4	Unsolved Total (667/13258)
DMTF1	Dominant Ultrarare	Infinite(13.20-infinite)	1.97 × 10^{-4}	2	0	Unsolved COU-NOS (187/12392)
BLVRB	Recessive Autosomal	Infinite(13.20-infinite)	1.97 × 10^{-4}	2	0	Unsolved COU-NOS (187/12392)
KLRD1	Dominant PTV	Infinite(9.38-infinite)	9.32 × 10^{-5}	3	0	Unsolved UPJO (276/10951)
EXOSC2	Dominant Ultrarare	Infinite(16.64-infinite)	1.86 × 10^{-5}	3	0	Unsolved UPJO (276/10951)
MMP15	Dominant Rare	16.29 (4.22-53.39)	6.86 × 10^{-5}	5	17	Unsolved UVJO (172/10093)

Table 1 List of the top candidate exome-wide suggestive signals. OR=odds ratio; PTV=protein truncating variant.

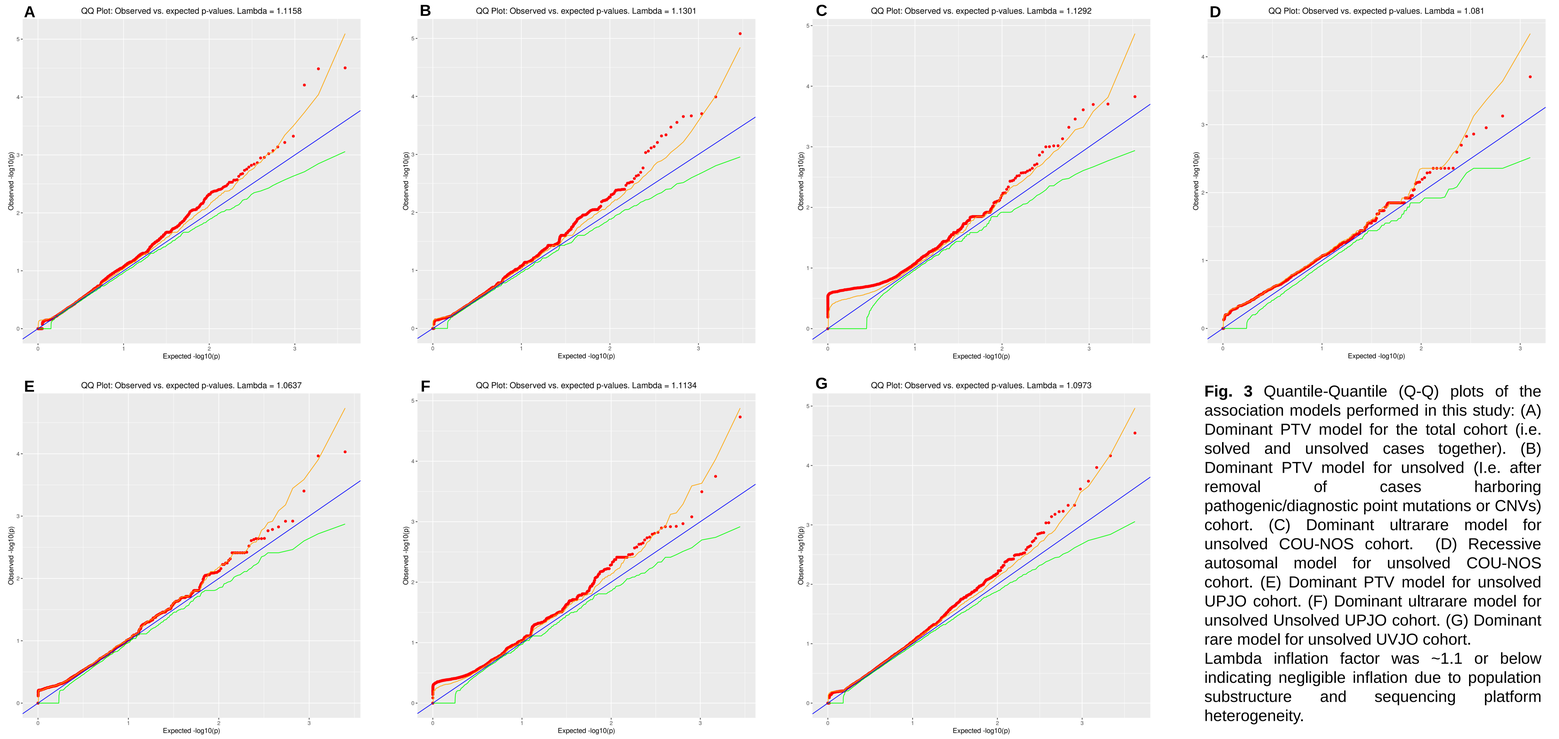

Fig. 3 Quantile-Quantile (Q-Q) plots of the association models performed in this study: (A) Dominant PTV model for the total cohort (i.e. solved and unsolved cases together). (B) Dominant PTV model for unsolved (I.e. after removal of cases harboring pathogenic/diagnostic point mutations or CNVs) cohort. (C) Dominant ultrarare model for unsolved COU-NOS cohort. (D) Recessive autosomal model for unsolved COU-NOS cohort. (E) Dominant PTV model for unsolved UPJO cohort. (F) Dominant ultrarare model for unsolved Unsolved UPJO cohort. (G) Dominant rare model for unsolved UVJO cohort.
Lambda inflation factor was ~1.1 or below indicating negligible inflation due to population substructure and sequencing platform heterogeneity.

Conclusions, Limitations, and Future Directions

- This study highlights the high genetic heterogeneity and complex architecture of COU. In fact, despite moderate to large sample size, no single genetic driver with large effect size could be identified by exome-wide collapsing analysis.
- Nevertheless, our association study identifies novel candidate genes for COU and related sub-phenotypes that, provided genetic and functional validation, will expand and deepen our understanding of the genetic underpinning of COU.
- Larger cohorts and/or integrated multidisciplinary approaches will be needed to robustly identify and validate novel genes predisposing to COU.

References

1. Thom RP et al. Pediatric nephrology (Berlin, Germany) 2013;28(9):1757-61
2. Ren Z et al. BMC Bioinformatics 2021 Mar 23;22(1):149
3. Subelj L et al. Physical review E, Statistical, nonlinear, and soft matter physics. 2011;83(3 Pt 2):036103
4. Mantel N et al. Journal of the National Cancer Institute. 1959;22(4):719-48
5. Richards S et al. Genetics in Medicine 2015 May;17(5):405-24
6. Ahmad N. Abou Tayoun et al. Hum Mutation 2018 Nov;39(11):1517-1524
7. Lim TY et al. Annual CAIRIBU Meeting Program 2021 Poster session 1, poster number 7

Funding Sources
NIDDK P20 DK116191,
RO1 DK103184, and R01 DK115574