Identifying Genetic Susceptibility Factors for Posterior Urethral Valves (PUV)

Yask Gupta1, Tze Y Lim1, Jonathan Diah1, Gina Jin1, Juntao Ke1, Martino Jeremiah1, Cathy L Mendelsohn2, David B Goldstein3, Simone Sanna-Cherchi1

1Division of Nephrology, Department of Medicine, Columbia University, New York, New York. 2Department of Urology, Columbia University, New York, NY. 3Institute for Genomic Medicine, Columbia University, New York, New York.

Background

Posterior urethral valves (PUV) indicate the presence of congenital obstruction at the level of the prostatic urethra (hence present only in males), and it is the most common congenital cause of lower urinary tract obstruction in males. Although affecting only 1 in 4,000 to 7,500 infants, PUV is responsible for up to 17% of pediatric end-stage kidney failure despite early surgical intervention. The role of genetic factors underlying PUV is unclear; the increased prevalence in patients with Down syndrome suggests that gene dosage might be a critical factor in the etiology of PUV, while the familial occurrence in rare cases suggests that single-gene defects can underlie PUV. Nevertheless, no mutations associated to isolated PUV have been identified so far, providing a unique opportunity for gene discovery.

Cohort Selection & Methods

- We performed a whole genome and exome sequencing study in 73 PUV trios (219 samples) and 17,865 controls. We then conducted two main gene discovery analyses: a) a case control analysis for enrichment of rare deleterious variants (exome-wide collapsing analysis); and b) a trio based analysis testing global and per-gene enrichment of ultra-rare de novo mutations.
- Exome-wide collapsing analysis (a.k.a. burden test): we used the 73 PUV index cases and compared them to 3,987 genetically-matched male controls (from 17,865 controls. We performed Principal Component Analysis (PCA) for dimensionality reduction on a set of predefined variants to capture population structure (Fig. 1a). Using the first 6 principal components (PCs) as input, we applied the Louvain method of community detection for identifying clusters within the data that reflect the ethnicity of the samples (Fig. 1b). We then selected high quality rare qualifying variants (QV) to perform collapsing analysis within each cluster. Afterwards, we extracted the number of candidates/cases and use the Cochran-Mantel-Haenszel (CMH) test to assess the association between case/control status and QV status while controlling for cluster membership.

PCA and Louvain Clustering of 73 cases and 17,865 controls

Results

Exome based collapsing (a.k.a. burden) analysis identifies three potential novel candidate genes for PUV

- Our clustering and harmonization methods allowed precise case-control matching and mitigation of sequencing platform heterogeneity as demonstrated by an inflation factor <1.1 in all models and similar distribution of synonymous variants in cases and controls (Fig. 2).
- Using exome-based collapsing methods, though we did not observe any study-wise significant association (P<1x10^-7); after correction for multiple models tested (Fig. 3). We detected one near-genome-wide significant signal PLCD1 (P=2.01x10^-6; OR=162) and two suggestive for DNAH11 (P=2.89x10^-6; OR=11) and TLK2 (P=5.99x10^-6; OR=104) as shown in Tab. 1.
- We next cross-referenced these findings with a mouse single-cell RNA-sequencing study on urethra, publicly available from Dr. Strand lab, and found that mouse orthologues Plcd1 was expressed in urethral luminal epithelial and Tk2 in all urethral epithelial cells.

Conclusions, Limitations, and Future Directions

- Here we provide the initial results and the optimization of the workflow for gene identification in PUV using trio-based and case-control analyses from whole-genome sequencing data.
- By analyzing the coding regions only of the genome (exomes), we report on several suggestive candidate genes for PUV that will need to be validate by genetic and functional studies.
- We have now expanded our whole genome sequencing study to over 150 PUV trios (more than double than the reported here) and acquired WGS data from >700 unaffected trios from the ClinVar database.
- We next used simulation-based statistical tests (as implemented in deNovoWest software) to explore the per-gene excess of de novo variants. While we did not observe excess of synonymous DMNs for any gene as a proof of concept, we did suggest effective excess burden (P<1x10^-5) for damaging de novo variants for the following genes: FOX2P2, MSLN, COL23A1, DRD4, SKA3, KRT77.

References

1. Sanna-Cherchi et al, J Clin Invest 2018; 128:7670-74

Funding

M01-RR00007, R01 DK011574, and R01 DK011574

Fig. 1. a) PCA showing distribution of cases and controls across all ancestries. b) UMAP clustering of our PUV cohort showing improved resolution of population ancestries.

Fig. 2 Histogram plot showing pre-individual number of synonymous variants in the largest cluster. The nearly equal distribution of variants in cases (green) and controls (blue) indicate good matching for downstream analyses.

Fig. 3 Manhattan plot of the CMH results from the ultra-rare model in the collapsing analysis. Each dot corresponds to a gene, reported in their genetic coordinates along the 23 chromosomes (x-axis). The blue line indicates the standard exome-wide significance threshold of 2.5 x 10^-8.

Fig. 4 The single cell RNA-seq (Strand Lab) shows expression of Plcd1, Tk2 and DNAH11 in mouse urethra cells. The umap plot shows data from single cell RNA-seq of both adult mouse (10-12 weeks). The clusters are identified as BE: basal epithelial, Ur: urethral luminal epithelial; Vp: VAP, ventral prostate luminal epithelium; DLPi: LE, dorsolateral prostate luminal epithelium; Ap: LE, anterior prostate luminal epithelium; Ed: Ed, anterior prostate luminal epithelium; F: Forsman, fibromuscular stroma; End, Endothelia; Leu, leukocytes.

Tab. 1. Summary results from the exome-wide burden of de novo variants using denovoanalyzer. We observed a significant enrichment of de novo loss-of-function mutations, while no difference in de novo synonymous variants in our cases as compared to expected, suggesting a role of de novo LOF in the pathogenesis of PUV.

Tab. 2. Summary results from the exome-wide burden of de novo variants using denovoanalyzer. We observed a -4-fold increase in burden of de novo LOF (P=2.8 x 10^-6), while no excess was found for either synonymous or missense variants (P>0.1).

Fig. 3. Deep sequencing of the urethra cell clusters. PLCD1, DNAH11 and TLK2 show enrichment of expression.