Highly sensitive variant detection in VNTR is possible with boosted exome sequencing: the example of MUC1-related nephropathy

Hassan **Saei**[1], Vincent **Morinière**[2], Laurence **Heidet**[1,3], Olivier **Gribouval**[1], Said **Lebbah**[2], Frederic **Tores**[4], Bertrand **Knebelmann**[5], Manon **Mautret-Godefroy**[2], Stephane **Burtey**[6], Vincent **Vuiblet**[7], Corinne **Antignac**[1,2], Patrick **Nitschké**[4], Guillaume **Dorval**[1,2]

[1] Labratory of hereditary kidney diseases, Inserm UMR 1163, Institut Imagine, Université Paris Cité, Paris, France [2] Medical genomics services for rare diseases, Hôpital Necker-Enfants Malades, Assistance publique, Hôpitaux de Paris (AP-HP), Paris, France [3] Pediatric Nephrology Service, Centre de Référence MARHEA, Hôpital Necker-Enfants Malades, Assistance publique, Hôpitaux de Paris (AP-HP), Paris, France

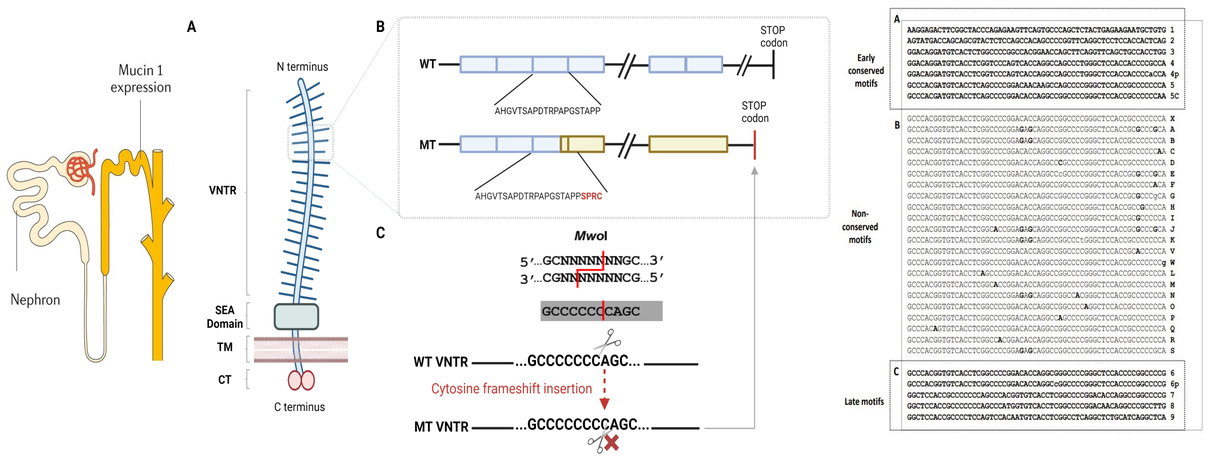
[4] Bioinformatic Platform, Inserm UMR 1163, Institut Imagine, Université Paris Cité, Paris, France [5] Nephrology Service, Centre de Référence MARHEA, Hôpital Necker-Enfants Malades, Assistance publique, Hôpitaux de Paris (AP-HP), Paris, France

[6] INSERM, C2VN, INRAE, C2VN, Aix-Marseille Université, Marseille, France. [7] Nephrology Service, CHU de Reims, Reims, France

Scan to read

INTRODUCTION

- Human genome comprises 3% of tandem repeats with variable length, a few of which have been linked to human rare diseases.
- Genotyping VNTRs using short-read sequencing data is challenging due to the poor read mappability.
- Autosomal dominant tubulointerstitial kidney disease-*MUC1* is caused by specific frameshift variants in the coding VNTR of the *MUC1* gene[1].
- *MUC1* encodes mucin-1 protein which is the main component of the mucus expressed in the distal tubules and collecting ducts of the nephrons.

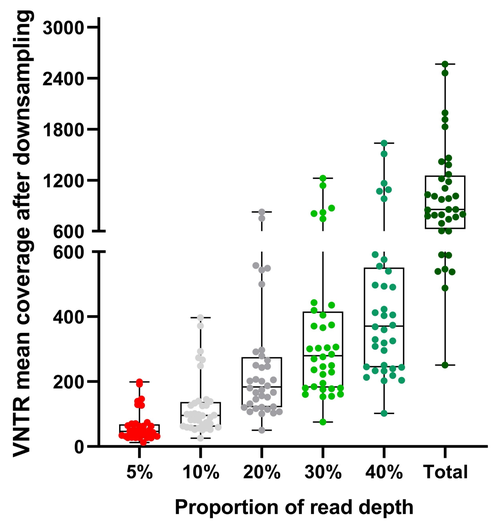


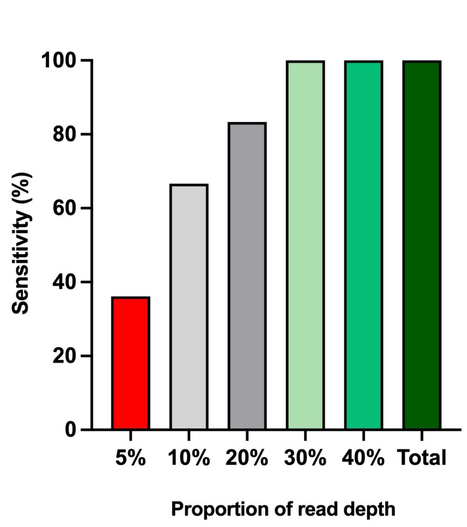
AIM

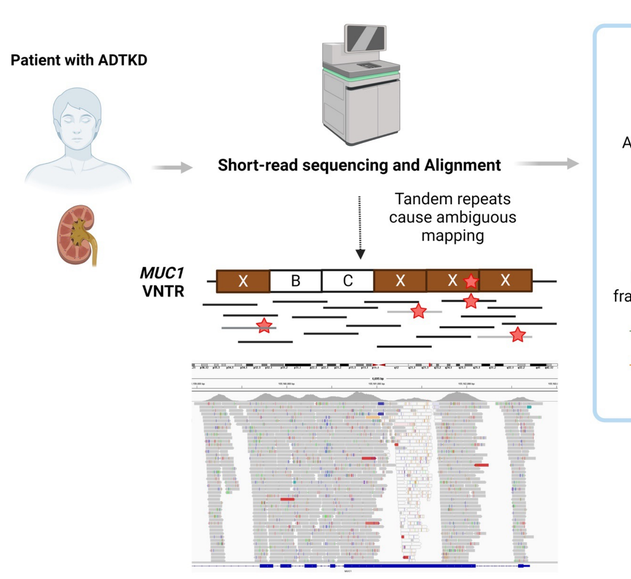
To enhance the genetic diagnosis and detection rate of ADTKD-*MUC1* by implementing standard short-read sequencing technology.

VNTR COVERAGE AND GENOTYPING SENSITIVITY

- The adequate VNTR sequencing depth is critical for highly sensitive genotyping.
- With downsampling alignment files from 60% to 95% of the total read depth, we studied the mean *MUC1* gene VNTR coverage. The mean coverage in targeted sequencing data was ~700x for the VNTR region.
- We studied the corrolation between mean VNTR coverage and genotyping sensitivity.
- We confirmed that the genotyping sensitivity has positive corrolation with the VNTR coverage. Below the coverage of 200x the sensitity drops below 90%.







Short reads originating from the VNTR region may map to multiple motifs and are filtered out during the variant calling step. To address this issue, we have developed a pipeline called VNtyper which uses K-mer frequencies for genotyping.

BOOSTING VNTR ENRICHMENT

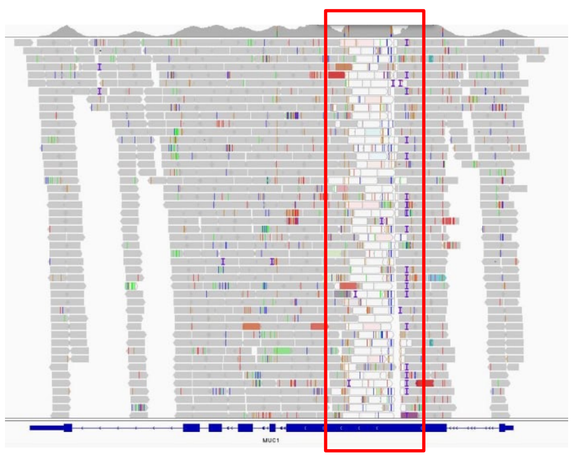
There are TWO main chalenges in genotyping VNTRs:

1. **Dilution effect**
2. **Capturing difference between library preparation technologies**

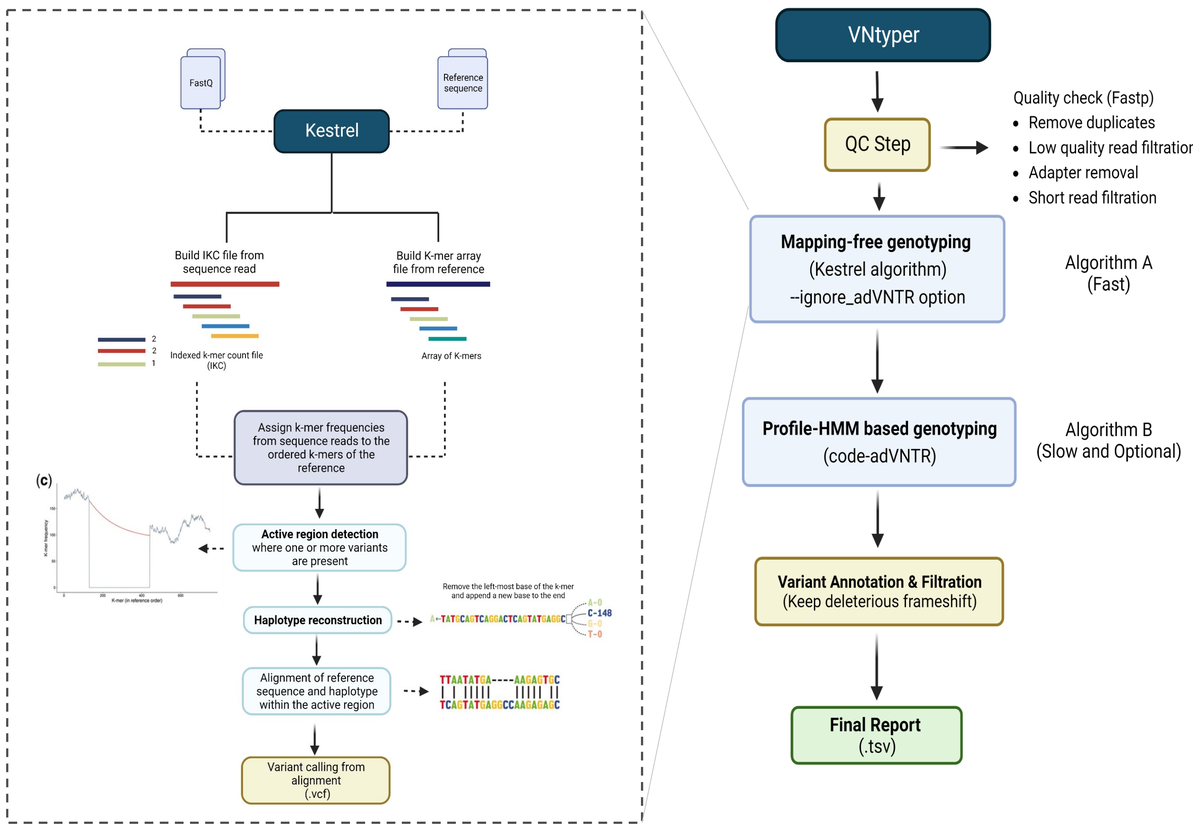
The dilution effects relate to the VNTR length (the number of times motifs are repeated), while the enrichment issue pertains to the capturing capacity of various library preparation technologies.

Regular exome **VNTR** region / Capturing technology A

Boosted exome **VNTR** region / Capturing technology B



HOW VNtyper WORKS?



VNtyper IMPROVES ADTKD DIAGNOSIS

- We applied VNtyper on **4040 patients** tested with hereditary renal disease panel and we identified **33 patients** with confirmed *MUC1* variation. All these cases were overlooked before this investigation.
- We picked 13 *ADTKD-MUC1* positive and three true negatives and performed exome sequencing with two different enrichment kits and with/without boosting the capture and compared the results.
- We discovered significant difference between target enrichment technologies in VNTR capturing. Boosting capture with spike-in probes could increases the genotyping sensitivity.
- In conclusion, VNtyper is designed on targeted sequencing data and could be applied on the exome data prepared with boosted or with an efficient enrichment technology.

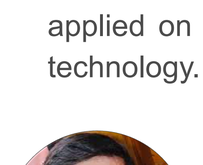
REFERENCES

[1] **Kirby et al**. Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in MUC1 missed by massively parallel sequencing. **Nature Genetics**, 2013.

[2] **Audano et al**. Mapping-free variant calling using haplotype reconstruction from k-mer frequencies. **Bioinformatics**, 2018.

[3] **Park et al**. Detecting tandem repeat variants in coding regions using code-adVNTR. **iScience**, 2022.

[4] **Saei et al**. VNtyper enables accurate alignment-free genotyping of MUC1 coding VNTR using short-read sequencing data in autosomal dominant tubulointerstitial kidney disease. **iScience**, 2023.



Hassan Saei

- PhD Candidate (PPU-Imagine International Doctoral Program fellow)
- I am passionate about leveraging computational methods to enhance genetic diagnosis. My profound interest lies in the development of disease models, such as organoids, and the application of genome editing techniques to delve into disease pathobiology and advance therapeutic solutions. @HassanSaei