Lifting the veil on Challenging Medically Relevant Genes

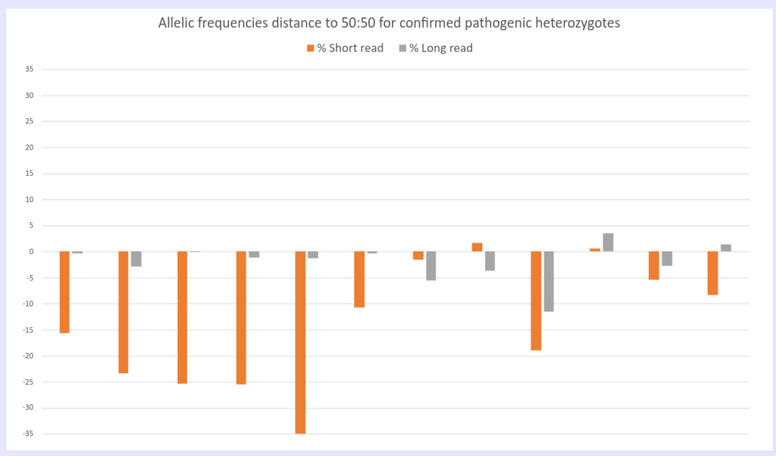
Background

Segmental duplications of genes of clinical interest have an negative impact on genetic diagnosis[1]. Indeed, the sequencing of these genes, either by Sanger or Short read methods, leads to partial, or incorrect characterization and identification of pathogenic variants. For this reason, we investigated the use of Long read sequencing for clinical settings. We created an in-house cohort of patients with Autosomal Dominant Polycystic Kidney Disease and sequenced them for the *PKD1* gene, which has multiple pseudogenes[2]. We also analyzed public data of the *FLG* gene, a gene with a Variable Number Tandem Repeat in its third exon, to investigate the use of Long read technologies for Atopic Dermatitis diagnosis[3].

PKD1

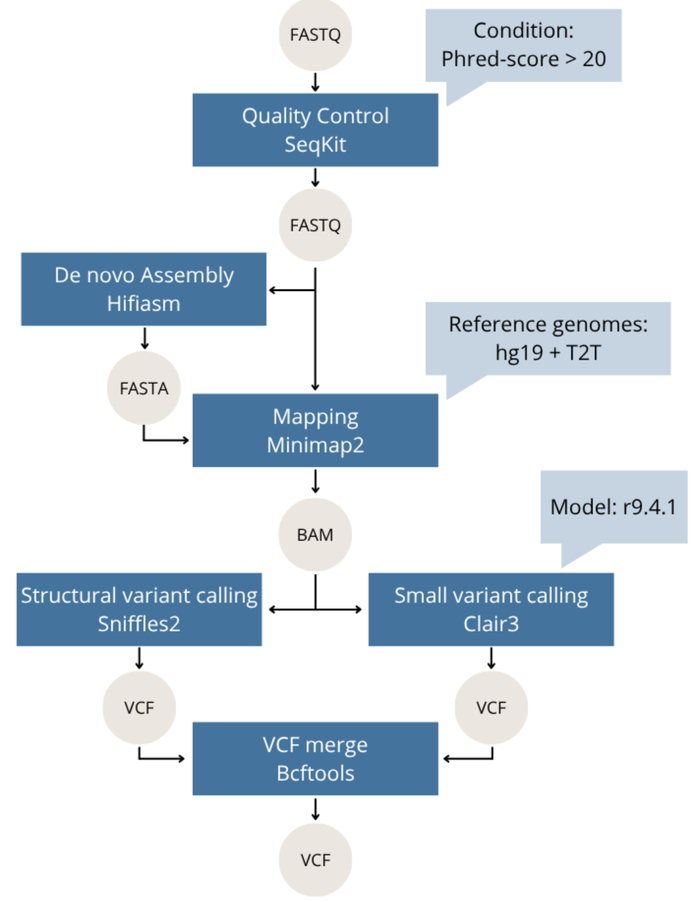
Patients	34
Confirmed by	Sanger
Variants found	56
Short read	Illumina Exome
Long read	ONT Amplicon
Variants retrieved	56/56

Pathogenic variants can be found with both Short read exome and Long read amplicon sequencing.



Long-read sequencing improves characterization of pathogenic variants.

Pipeline

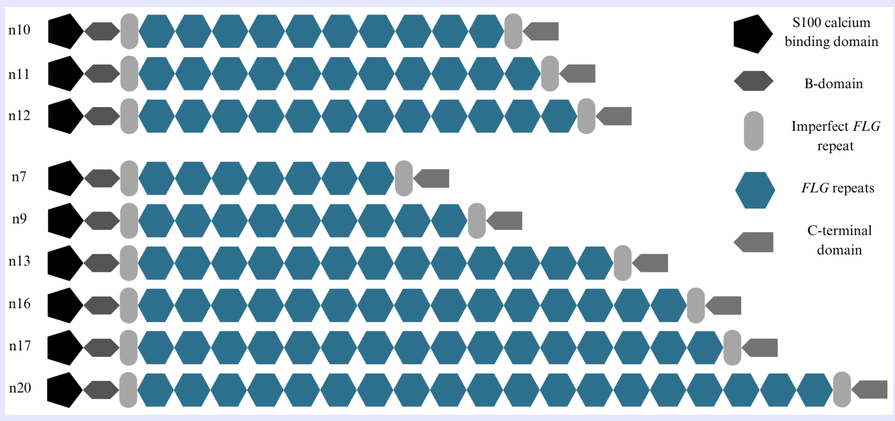


Nextflow pipeline using Apptainer containerized Long read analysis tools.

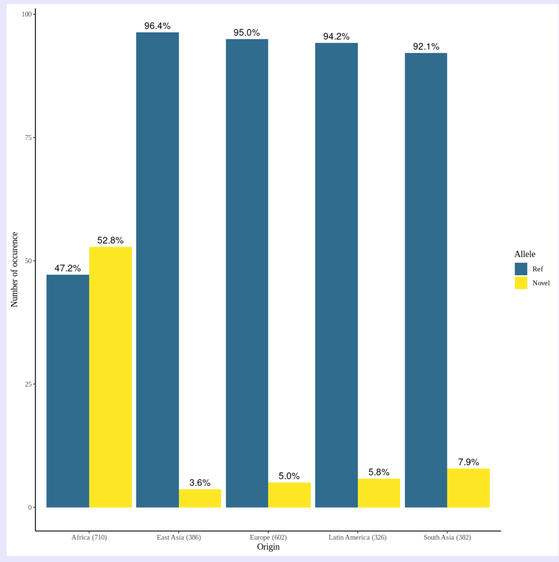
FLG

Region	Populations	Samples	X≥ 2
Africa	7	316	279
Europe	7	200	185
East Asia	5	213	193
South Asia	5	223	191
Latin America	4	182	163

We analyzed 1011 Long read public samples 28 different origins.



Novel *FLG* alleles discovered, with unusual number of repetitions[4].



Novel *FLG* alleles are mostly found in African populations.

Relevance

Long read sequencing improves the characterization of the *PKD1* and *FLG* genes, by bypassing the limitation of Short read technologies. While investigating *FLG* data, novel alleles of this gene, never described before, were found. We believe that they may originate from non-allelic homologous recombination. These novel allele could impact Atopic Dermatitis, as the number of repetition may have an impact on the severity of the disease[5]. Other challenging genes of clinical interest, such as *MUC1* could also benefit from Long read sequencing, however the reference genome is a limitation for correct characterization, and we believe that aligning data to a pangenome of this gene could improve results.

Contact me!

References

1 Mandelker D, et al. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. Genetics in Medicine. déc 2016;18(12):1282-9.

2 Lanktree MB, Haghighi A, et al. Insights into Autosomal Dominant Polycystic Kidney Disease from Genetic Studies. CJASN 2021;16:790–9.

3 Osawa R, et al. Filaggrin Gene Defects and the Risk of Developing Allergic Disorders. Allergology International. 2011;60(1):1-9.

4 Eaaswarkhanth M, et al. Atopic Dermatitis Susceptibility Variants in Filaggrin Hitchhike Hornerin Selective Sweep. Genome Biol Evol. oct 2016;8(10):3240-55.

5 Brown SJ, et al. Intragenic Copy Number Variation within Filaggrin Contributes to the Risk of Atopic Dermatitis with a Dose-Dependent Effect. Journal of Investigative Dermatology 2012;132:98–104.

Victor Grentzinger[1,2], Leonor Palmeira[2], Keith Durkin[1,2], Maria Artesi[1,2], Vincent Bours[1,2]

victor.grentzinger@uliege.be

1 **GIGA - Université de Liège,** 2 **CHU de Liège**