Lifting the veil on Challenging Medically Relevant Genes

Victor Grentzinger

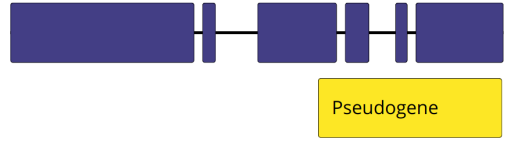
Laboratory of Human Genetics
GIGA Institute - Liège, Belgium

SCS 20/07/25

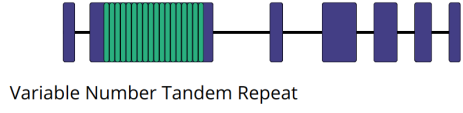
What is a Challenging Medically Relevant Gene?

Challenging genes

Highly similar to other DNA sequences within the same genome.



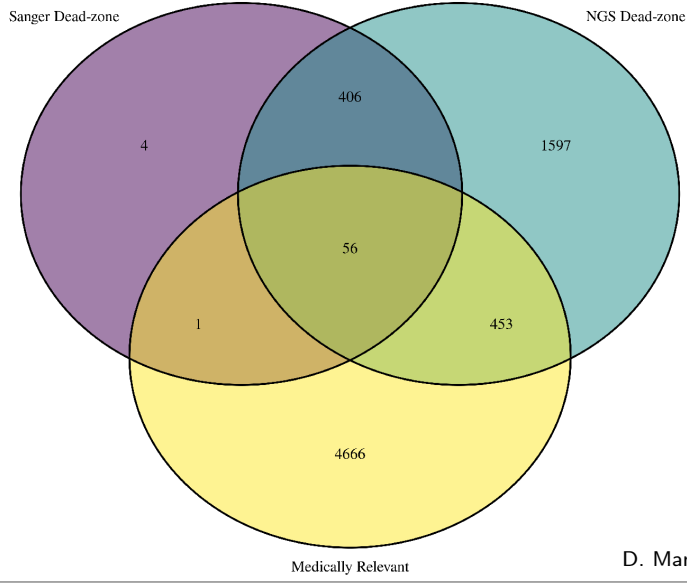
These homologous regions are really complex to characterize.



Two major categories:

- Sanger deadzone.
- NGS deadzone.

Challenging Medically Relevant Genes (CMRG)



D. Mandelker et al. (2016)

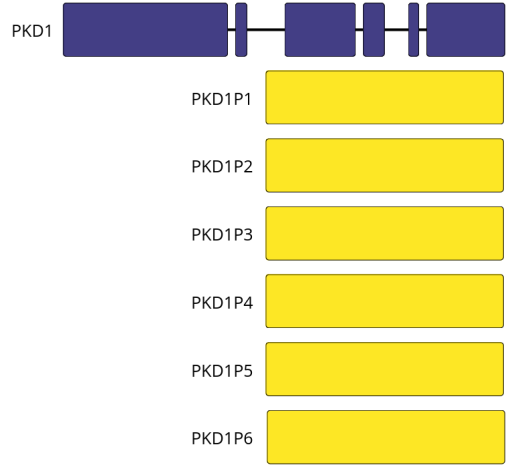
My research goals

- Improve our knowledge about the sequence structure of these challenging genes.
- Understand how can we overcome the NGS limitation, mostly using long-read sequencing.
- Fast and cost-effective methods of genetic diagnosis.

Genes of interest

- *PKD1*: Autosomal Dominant Polycystic Kidney Disease (ADPKD).
- *FLG*: Atopic Dermatitis (AD).

The *PKD1* case



Responsible for 75% of Autosomal Dominant Polycystic Kidney Disease.

6 pseudogenes that are highly similar

Part of the NGS deadzone.

PKD1 strategy

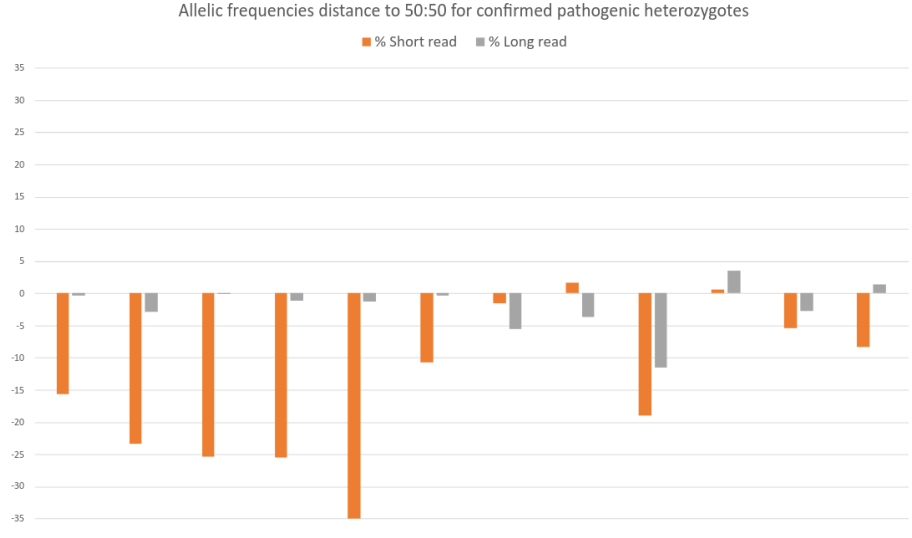
Gold standard: Sanger sequencing.

Dataset of 34 patients with 56 pathogenic variants, sequenced both with:

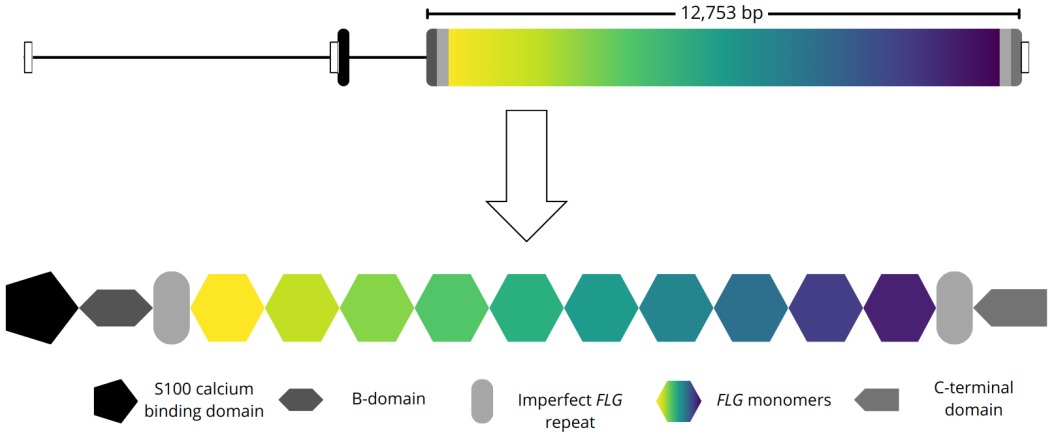
- Whole Exome Sequencing with Illumina.
- Amplicon sequencing with Oxford Nanopore Technologies.

All variants have been found with *both* techniques.

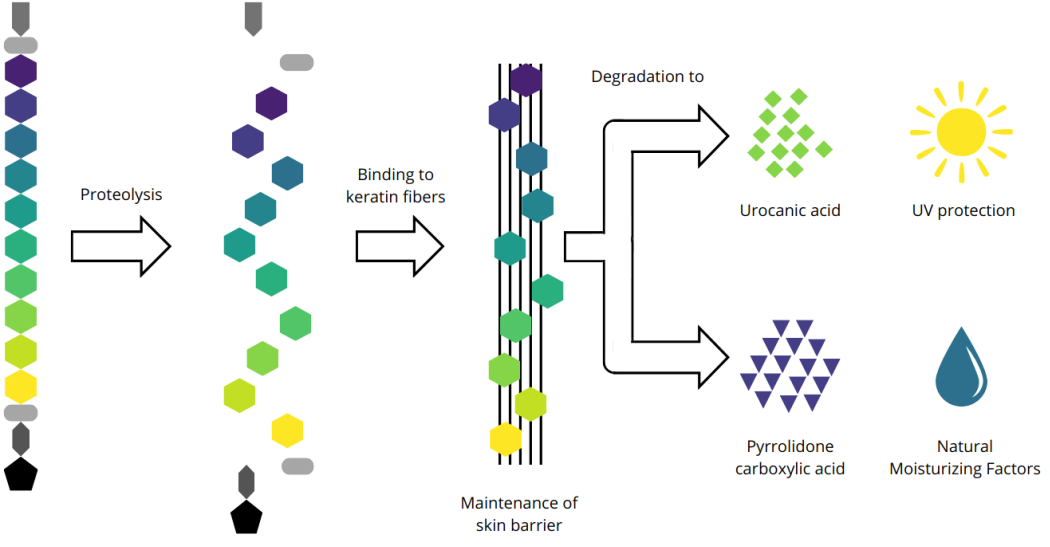
Long-read improves allele frequency



The Filaggrin gene (*FLG*)



Role of *FLG*

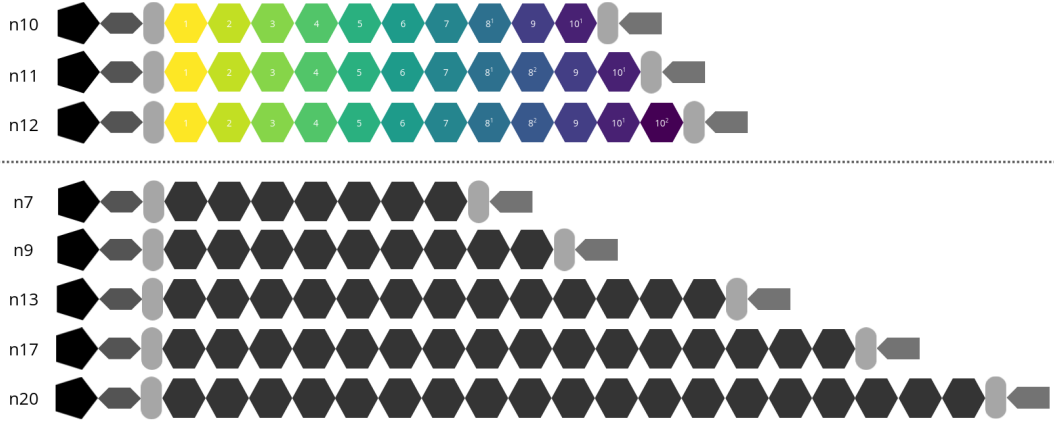

Proteolysis
Binding to
keratin fibers
Degradation to
Urocanic acid
UV protection
Pyrrolidone
carboxylic acid
Natural
Moisturizing Factors
Maintenance of
skin barrier

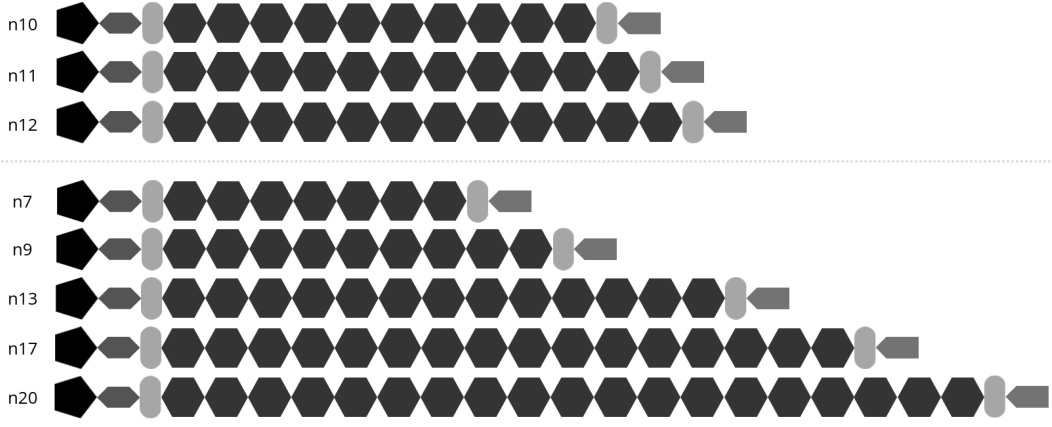
Multiple known alleles of *FLG*

Public multi-ethnic cohort

Origin	Number of samples	After QC
Africa	316	279
Europe	200	185
East Asia	213	193
South Asia	223	191
Latin America	182	163
Total	1134	1011

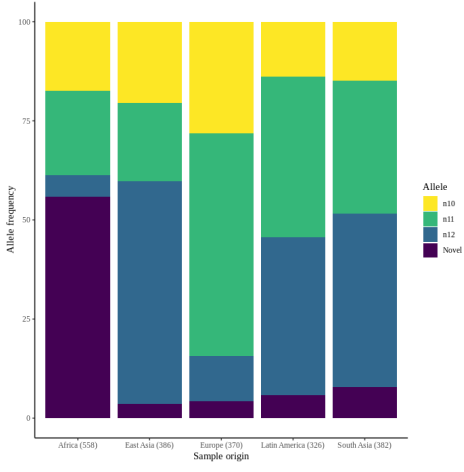
Catalog of novel alleles




n10
n11
n12
n7
n9
n13
n17
n20

Frequencies of novel alleles

Allele	Africa	Europe	East Asia	South Asia	Latin America
n10	97	104	79	57	45
n11	119	**208**	76	128	**132**
n12	30	42	***217***	**167**	130
n7	0	0	1	0	0
n9	45	1	0	0	0
n10	***228***	3	4	3	7
n11	5	9	9	26	9
n12	6	2	0	1	1
n13	28	0	0	0	1
n17	0	1	0	0	0
n20	0	0	0	0	1



Potential origin of these novel alleles

- Seems to originate from misalignment during homologous recombination.
- Due to the high similarity between repeats, an misalignment could happen during DNA repair.

Conclusion

- Long-read sequencing can improve the characterization of CMRG, and even be used in a clinical setting.
- Novel *FLG* alleles have been identified, mostly in African and African-descent populations.
- However, some regions are still difficult to characterize.
  - Nanopore adaptive sampling or a pangenome reference could help.

Acknowledgement

Pr. Vincent Bours
Dr. Keith Durkin
Dr. Leonor Palmeira
Dr. Maria Artesi
Dr. Benoit Charloteaux
Dr. Laura Helou
Pr. Vinciane Dideberg
Nadine Cambisano
Nathalie Renotte
Dr. Sabine Olivier
Dr. Annette Uwineza
Dr. Alice Amani Uwajeni
Pr. Alan Irvine

Contact me!