Genomic insights into Parkinson's disease of a large Italian cohort

Fant A.[1], Musacchia F.[1], Trova S.[1], Gaudio A.[2], Landuzzi F.[1], Treves G.[1], Pandolfini L.[4], Monfrini E.[5], Di Fonzo A.[6], Trevisan L.[3], Cavalli A.[1], Di Giovanni M.[7], Lillia N.[7], Caligiana L.[7], Cordera S.[7], Giardini G.[7], Vecchi M.[1], Mandich P.[2], Gustincich S.[1]

[1] Istituto Italiano di Tecnologia, CMP3VDA, Aosta, Italy; [2] IRCCS Ospedale Policlinico San Martino, UOC Genetica Medica, Genova, Italy; [3] University of Genoa, Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics and Maternal and Child Health, Genova, Italy; [4] Istituto Italiano di Tecnologia, Center for Human Technologies (CHT), Genova, Italy; [5] Dino Ferrari Center, Neuroscience Section, Department of Pathophysiology and Transplantation University of Milan, Milano, Italy; [6] Foundation IRCCS Ca' Granda Ospedale Maggiore Policlinico, Neurology Unit, Milano, Italy; [7] Neurologia e Stroke Unit, Ospedale Regionale "Umberto Parini", Aosta, Italy

1. BACKGROUND

The 5000Genomi@VDA research project, which started on Dec 01, 2019 in Aosta, aims to sequence 5,000 genomes of subjects in close collaboration with Aosta Regional Hospital "Umberto Parini" and research centers and hospitals in northern Italy. Genomic analysis of neurodegenerative disease, particularly Parkinson's disease (PD), is one of the 5 sub-projects covered within the 5000Genomi@VDA project.

2. METHODS

2.1 Cohort

We have performed Whole Genome Sequencing with short-reads (sr-WGS) NovaSeq 6000 Illumina platform on 450 subjects and long-read Nanopore sequencing (lr-WGS) in selected cases.

	N	Sex (%male)	Mean age	Onset
PD	300	58%	62	37% EOPD
Healthy	150	64%	59	/

EOPD: Early onset Parkinson's disease (< 50yo)

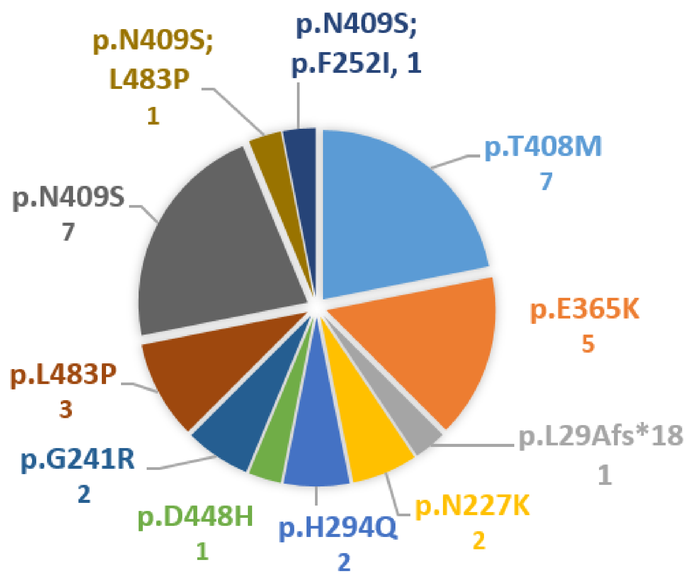
2.2 Pipeline

Bioinformatic analyses were conducted using NVIDIA Clara™ Parabricks®, which is a GPU-accelerated suite comprising pre-built pipelines and standalone tools from the GATK suite. This enabled the generation of a VCF file, from FASTQ data, in less than an hour. For SV detection, in addition to CNVkit, we incorporated a consensus pipeline, that integrates 4 different tools (Manta, BreakDancer, Lumpy, CNVnator) to enhance accuracy by reducing false positive calls.

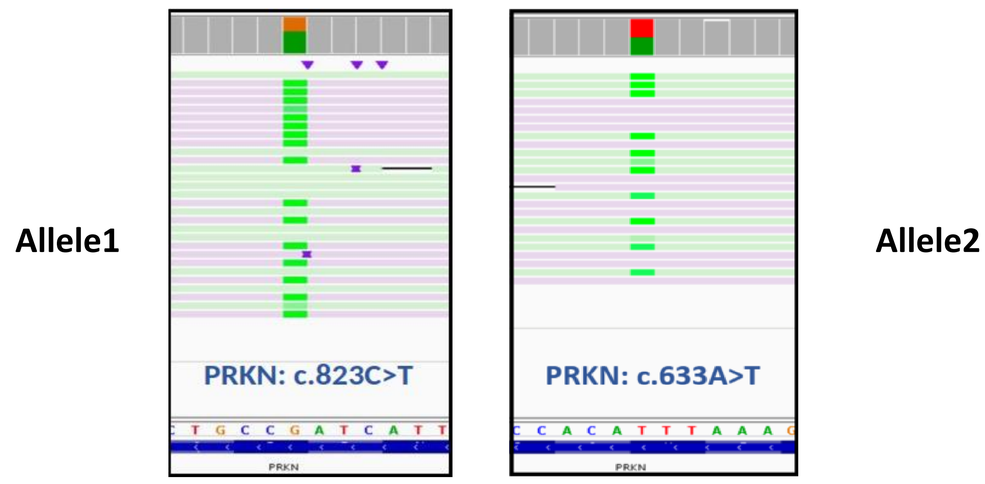
3. RESULTS

3.1 Parkinson's Disease known variants in GBA

Within our cohort of PD patients, approximately 10% (32/300) possess a GBA mutation. This aligns with existing literature suggesting that GBA mutations are present in 5% to 15% of PD cases (Smith L. et al, Cells, 2022).

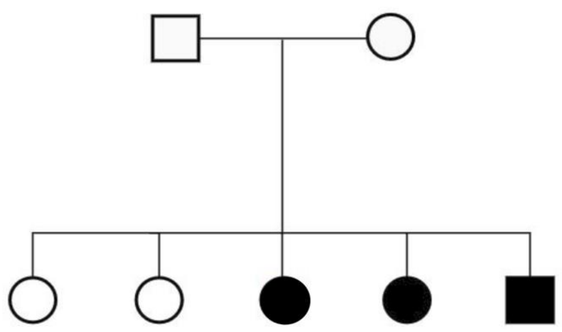


3.2 Compound heterozygosity of SNV

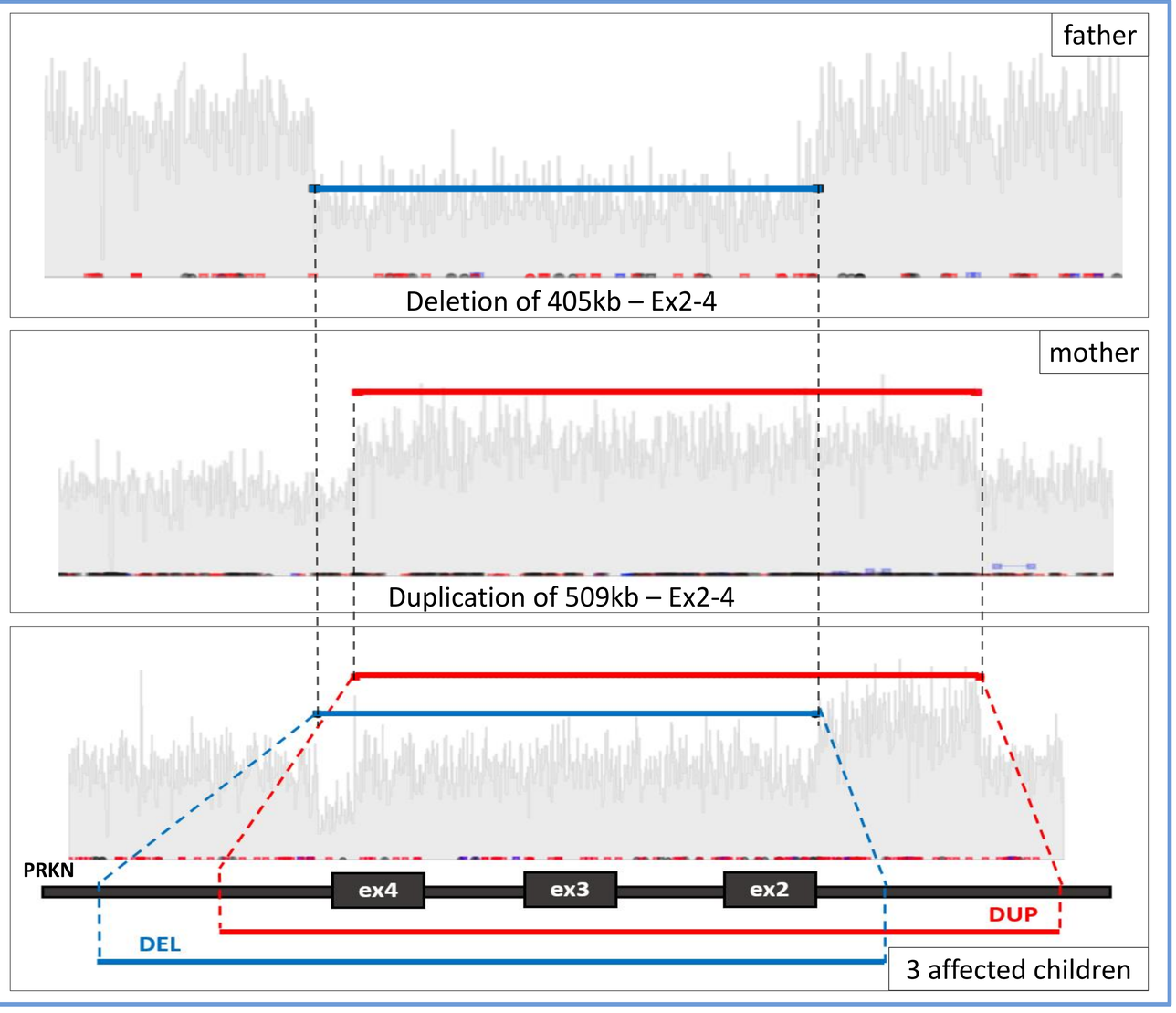


With sr-WGS, we identified one patient harboring two pathogenic variants in the PRKN gene. By employing lr-WGS, we were able to resolve phasing and confirm compound heterozygosity.

3.3 Compound heterozygosity of CNV



Interestingly, in 3 PD affected siblings from the same family we identified a simultaneous **deletion** on one allele and **duplication** on the other allele within the same DNA region of *PRKN*, which, to our knowledge, has not been described before. CNVs impacting the initial exons of the 1.38 Mb *PRKN* gene are common. However, CNs involving Ex2-4 display a low occurrence rate (E2-4del: 1,4% in UK BioBank, 1,6% in NIH-PD/AMP-PD, 0% in gnomAD-SV v2.1; E2-4dup: 5% in UK BioBank, 0% in NIH-PD/AMP-PD, 3,5% in gnomAD-SV v2.1) *(Zhu W. et al., Brain, 2022).*



4. CONCLUSIONS and FUTURE PERSPECTIVES

WGS enabled the identification of diagnostic variants that conventional methods failed to identify: two distinct types of compound heterozygosity have been solved. Nanopore permitted the physical phasing of SNVs that were too distant to be resolved by sr-WGS, while sr-WGS revealed CNVs, not detected using MLPA (Multiplex Ligation-dependent Probe Amplification). In addition, we are focusing on non-coding regions, analyzing variants in enhancer/promoter that may impact gene function and/or expression in this Italian cohort.