Huntington disease (HD) is a dominantly inherited neurodegenerative disorder caused by the expansion of the CAG repeat in the huntingtin (*HTT*) gene. The clinical phenotype of HD is commonly described as a triad of movement disorder, cognitive decline and psychiatric complications. In South Africa, patients with Africa ancestry presenting with an HD phenotype may alternatively have Huntington disease like-2 (HDL2), caused by the expansion of the CTG repeat in the junctophilin-3 (*JPH3*) gene. Although the underlying pathological mechanism in HD/HDL2 remains unclear, the main driver of disease severity is the expanded repeat. However, considerable additional variation is observed that is heritable. Although HD has been extensively studied in patients with European ancestry, very few studies have been dedicated to studying patients with African ancestry with even fewer studies conducted on HDL2. This study aimed to characterise the diversity at the two loci responsible for HD and HDL2 and explore potential genetic modifiers of disease.

In the *HTT* analysis, a high level of sequence diversity was observed in African ancestry individuals with the identification of three novel allele structures. The most common disease-associated allele structure Q^1-2-0-9-2 was defined as atypical, due to the loss of the CCGCCA sequence. Although the CAG repeat was confirmed to be the main driver of disease severity, the Q^1-2-0-9-2 allele structure was associated with an earlier age of diagnosis of approximately 7.1 years and occurred exclusively on haplotype B2. Although the somatic expansion was associated with an earlier age of diagnosis, the Q^1-2-0-9-2 allele structure displayed reduced somatic expansion. Thus, the Q^1-2-0-9-2 allele structure occurring on haplotype B2 was proposed as an African *cis*-acting modifier that appears to modify the age of diagnosis, through a mechanism that was not driven by somatic expansion. This was a particularly interesting finding, as it was hypothesised that these patients would have the highest amount of somatic expansion however the inverse was found to be true. These findings provide a different perspective on HD and emphasize the need for the development of African inclusive therapeutic approaches.

In the *JPH3* analysis, less sequence diversity was observed across all individuals affected and unaffected with HDL2. The disease alleles all had the same typical allele structure, which occurred on a single disease haplotype 2: CCATC. The haplotype 2 was also only identified in African ancestry non-disease alleles lending support of an African specific single origin of the HDL2 mutation. The CTG repeat was confirmed to be the main driver of disease severity although when compared with the CAG repeat, it accounted for considerably less of the variation in the disease phenotype. When the broader haplotype background was considered,

the larger haplotype 6: CCCATCG was associated with an earlier age of onset of disease by 17.2 years however the outer two tag-SNPs appear less informative, and its mechanism is uncertain. The CTG repeat showed considerably less somatic expansion than the CAG repeat and unlike HD, somatic expansion was not significantly associated with the age of onset in HDL2 although the sample size of 28 HDL2 patient was smaller than the 60 HD patients. A major limitation when studying rare diseases is the sample size which inevitably reduces the power to identify significant associations.

The comparison of the *HTT* and *JPH3* loci indicated that despite the clinical similarities between HD and HDL2, stark contrasts were identified that suggest a different potential mechanism of disease modification may be occurring in the two diseases. With the only similarity seen in the association with the disease-associated repeats, making the clinical similarity more intriguing. This study has emphasised the importance of studying genetically diverse populations and shown the locus heterogeneity within the *HTT* locus, the disease phenotype and highlighted differences among an HD phenocopy.