



# Tracing the mutated *HTT* and haplotype of the African ancestor who spread Huntington disease into the Middle East

Ferdinando Squitieri, MD, PhD<sup>1</sup>, Tommaso Mazza, PhD<sup>2</sup>, Sabrina Maffi, PsyD<sup>1</sup>, Alessandro De Luca, PhD<sup>3</sup>, Qasem AlSalmi, MD<sup>4</sup>, Salma AlHarasi, PhD<sup>4</sup>, Jennifer A. Collins, BA<sup>5</sup>, Chris Kay, PhD<sup>5</sup>, Fiona Baine-Savanhu, PhD<sup>6</sup>, Bernard G. Landwhermeyer, MD<sup>7</sup>, Umberto Sabatini, MD<sup>8,9</sup> and Michael R. Hayden, MD<sup>5</sup>

**Purpose:** We aimed to determine the origin and genetic characteristics of Huntington disease (HD) in the Middle East.

**Methods:** We performed genetic and genealogical analyses to establish the ancestral origin of the *HTT* pathogenic variant from a large kindred from Oman (hereafter called the OM-HD-01 pedigree) by single-nucleotide polymorphism and dense haplotype analysis genotyping.

**Results:** We traced the oldest ancestry of the largest, eight-generation, OM-HD-01 pedigree ( $n = 302$  subjects, with 54 showing manifest HD) back to sub-Saharan Africa and identified a unique HD haplotype carried by all pedigree members, which consisted of portions of the C6 and C9 haplotypes and was carried by all affected members. Such a unique HD haplotype was of African origin and appeared to be associated with large CAG repeat expansions on average and high frequency of juvenile-onset HD.

Three other families from the same area were also identified and found carrying a Caucasian HD haplotype A, also shared by most families of Arab ancestry.

**Conclusion:** Mutated *HTT* spread into Middle East with a unique haplotype of African origin, appeared to be associated with juvenile-onset, a HD condition frequently occurring in Black Africans, and may have a significant impact on further development of novel targeted genetic therapies.

*Genetics in Medicine* (2020) <https://doi.org/10.1038/s41436-020-0895-1>

**Keywords:** Huntington disease; juvenile-onset Huntington disease; haplotype analysis; SNP; *HTT*

## INTRODUCTION

After the first concise disease description in 1872, the history, etiology, and clinical study of Huntington disease (HD, OMIM nb: 143100) have focused mostly on Caucasian populations of European ancestry, which show the highest rate of de novo expanded CAG repeat alleles in the huntingtin gene (*HTT*), associated with increased disease prevalence.<sup>1</sup>

In contrast, very little is known about HD in non-European, Black African ancestry populations, although evidence does suggest that when HD has occurred in these groups (e.g., Tanzania, Zimbabwe, South Africa, and Peru), it was either introduced by Europeans or originated from local pathogenic variants in *HTT*, as confirmed by haplotype analysis.<sup>2,3</sup> Haplotype analysis enables the population from which an *HTT* pathogenic variant has arisen to be determined because in European and Caucasian populations, *HTT* CAG expansions are typically found on haplogroup A, whereas in Black

African and East Asian populations, they are more commonly found on specific variants of haplogroups B and C.<sup>3–6</sup> Cases of HD in the Middle East have also been reported only sporadically and the ancestral origin of HD expanded alleles in these individuals, as well as the prevalence of the disease, remains unknown.<sup>7</sup>

In July 2013 we were contacted by a family living in Oman who reported having been affected by HD over several generations. Indeed, the family represents a large HD cluster (hereafter named OM-HD-01 pedigree), owing to its high number of intermarriages and subsequent increase in heritability risk. Given the gaps in our knowledge about HD in the Middle East, we took the opportunity to perform a study to determine the ancestral origin of *HTT* expanded alleles and haplotype from such large kindred, to determine genetic origin of HD in affected individuals in the Sultanate of Oman, the heart of the Middle East, and compare genetic

<sup>1</sup>Huntington and Rare Diseases Unit, Fondazione IRCCS Casa Sollievo della Sofferenza Research Hospital, San Giovanni Rotondo, Italy; <sup>2</sup>Bioinformatics Unit, Fondazione IRCCS Casa Sollievo della Sofferenza Research Hospital, San Giovanni Rotondo, Italy; <sup>3</sup>Medical Genetics Unit, Fondazione IRCCS Casa Sollievo della Sofferenza Research Hospital, San Giovanni Rotondo, Italy; <sup>4</sup>National Genetic Centre, Royal Hospital, Ministry of Health, Muscat, Sultanate of Oman; <sup>5</sup>Centre for Molecular Medicine Therapeutics, Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; <sup>6</sup>NHLS & School of Pathology, University of the Witwatersrand, Johannesburg, South Africa; <sup>7</sup>Abteilung Neurologie, Universitätsklinik Ulm, Ulm, Germany; <sup>8</sup>Department of Neuroradiology, University Magna Graecia, Catanzaro, Italy; <sup>9</sup>Italian League for Research on Huntington Disease (LIRH) Foundation, Rome, Italy. Correspondence: Ferdinando Squitieri (f.squitieri@css-mendel.it)

Submitted 15 April 2020; revised 24 June 2020; accepted: 24 June 2020

Published online: 14 July 2020

findings with what is currently known about other Arab and Caucasian HD populations.

## MATERIALS AND METHODS

### Ethics statement

All individuals provided written informed consent prior to participation in the study and in accordance with the Declaration of Helsinki. For participants who lacked capacity to consent, study sites followed country-specific guidelines for signing consent forms. Minors agreed with both parents authorizing for them. All authors were responsible for data collection and analyses; all authors had full access to the data, and all authors can vouch for the accuracy and completeness of the data and analyses. The study was ethically approved by the local institutional review boards of each institution (H05-70532—Genetic Modifiers of HD, University of British Columbia, Vancouver, Canada; protocol 1.050219/mob—LIRH Foundation and CSS-Mendel Institute, Rome, Italy; SRC 39/2020—Royal Hospital, Ministry of Health, Muscat, Sultanate of Oman).

Our study was conducted between August 2013 and March 2019. In the OM-HD-01 pedigree, a full genealogical analysis was performed to determine the ancestral origin of the HD pathogenic variant (Fig. 1). This information was gathered retrospectively via one-to-one interviews with all surviving members. DNA samples from patients in the OM-HD-01 pedigree were collected at the National Genetic Centre (NGC), Sultanate of Oman, the Ministry of Health Country Hospital, for both research and diagnostic purposes. Genetic analyses to determine CAG repeat length in the *HTT* gene were performed in parallel at the NGC and at the CSS-Mendel Institute in Rome, Italy, as described.<sup>8</sup> Single-nucleotide polymorphism (SNP) genotyping was performed at the Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada.<sup>9</sup> To determine the origin of the expanded allele of the OM-HD-01 pedigree, three selected HD individuals of this family underwent dense haplotype analysis, as previously described.<sup>9</sup> Follow-up haplotyping of a further 22 DNA samples (19 HD and three unaffected controls), from five related kindreds was subsequently performed to determine whether members of the extended OM-HD-01 pedigree (Figure S1) shared the same mutant *HTT* haplotype as these three original family members. For comparison, SNP analysis was also performed in three other unrelated Omani parent–offspring trios from pedigrees OM-HD-02, OM-HD-03, and OM-HD-04 (Figure S2).

Age at HD onset was recorded based on first relevant neurological symptoms alone or in association with psychiatric symptoms and according to either expert clinician assessment (affected alive patients) or retrospective recall via interviews with caregivers/family members (affected deceased patients).<sup>8</sup> The proportion of patients with juvenile-onset HD (joHD; age of neurological onset <20 years) was also determined.<sup>8</sup> We also included a reference cohort comprising European HD patients of Italian ancestry, who were either

enrolled into the Registry (NCT01590589; start date: June 2004, no longer active; [https://www.enroll-hd.org/enrollhd\\_documents/2016-10-R1/registry-protocol-3.0.pdf](https://www.enroll-hd.org/enrollhd_documents/2016-10-R1/registry-protocol-3.0.pdf)), ENROLL-HD (NCT01574053; start date: July 2012, still active; [https://www.enroll-hd.org/enrollhd\\_documents/Enroll-HD-Protocol-1.0.pdf](https://www.enroll-hd.org/enrollhd_documents/Enroll-HD-Protocol-1.0.pdf)), or LIRH Foundation institutional database (start date March 2001, still active).<sup>8</sup>

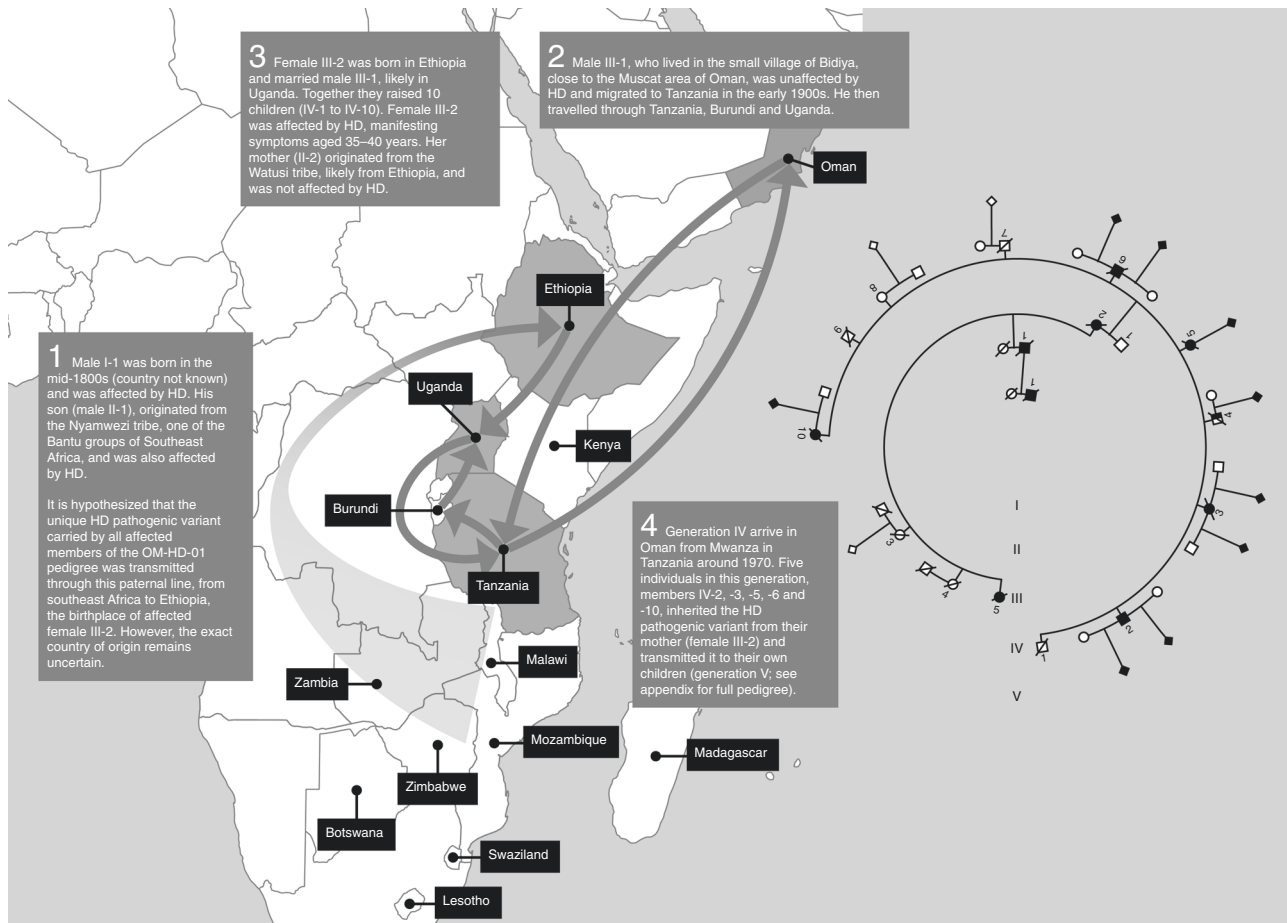
To compare the mean expanded repeat size and age of onset between patients in the Omani and European-Italian cohorts, patients with  $\geq 80$  CAG repeats and with  $< 40$  CAG repeats in *HTT* gene were excluded to minimize the bias due to the highly expanded or low penetrance allele size. Differences in clinical presentation between Omani and European-Italian patients were compared using Fisher's exact tests for categorical variables and Mann–Whitney *U* test for continuous variables. Normality distribution of data was verified by Shapiro–Wilk test. We used linear regression analysis after logarithm transformation to correlate age at onset with expanded CAG repeat length. When the *P* value for the groups was  $< 0.05$ , we concluded that there was significant difference between groups. All statistical analyses were performed in R 3.4.0.

## RESULTS

In the OM-HD-01 pedigree ( $n = 302$ ), the largest family spanning eight generations, 54 individuals had manifest HD, 8 of them with joHD (8/54 or 14.8%); of these 54 individuals, 33 were still alive, and 21 had been affected by HD but were now deceased. In addition, a further 241 were at risk (Figure S1). Genealogical analysis revealed that the oldest OM-HD-01 affected individual (born in the mid-1800s) had originated from sub-Saharan Africa, presumably from Southeast Africa, then moved to Ethiopia (Fig. 1).

Follow-up high density SNP genotyping of 22 samples from the extended family branches confirmed that a single unique C6x9 haplotype was carried by all examined individuals on the expanded HD allele (Fig. 2, Table S1). Table S2 shows each *HTT* allele phased to CAG repeat sizes and haplotypes of the three individuals selected from the pedigree. Careful review of our global haplotype data set identified an additional single Black South African individual sharing the same distinctive HD haplotype of the original OM-HD-01 family's HD ancestor. Further analysis revealed it to be a possible recombinant haplotype, consisting of portions of the C6 and C9 haplotypes (Fig. 2a, Table S1A), with the CAG repeat sequence situated in the C6 haplotype spanning the first half of the gene and the C9 haplotype spanning the distal half of the gene. Conversely, three unrelated Omani parent–offspring trios from unrelated pedigrees OM-HD-02, OM-HD-03, and OM-HD-04 were shown to have the HD expansion on the A2b haplotype (Fig. 2b, Table S1B).

Demographic and clinical patients' characteristics are included in Table S2. Mean age at neurologic disease onset was significantly lower in the OM-HD-01 pedigree than in the European-Italian cohort (34.5 vs. 43.8 years, respectively;  $p < 0.0001$ ) and mean CAG repeat length was significantly longer



**Fig. 1 Geographical ancestral migratory pathways between sub-Saharan Africa, and Oman, the Middle East.** Historical notes referred to OM-HD-01 individuals, illustrating the origin of the unique C6x9 African *HTT* haplotype in this pedigree. A draft of the first five OM-HD-01 pedigree generations is reported on the right side of the figure to allow interpretation of the migratory pathways between sub-Saharan Africa and Oman, and vice versa.

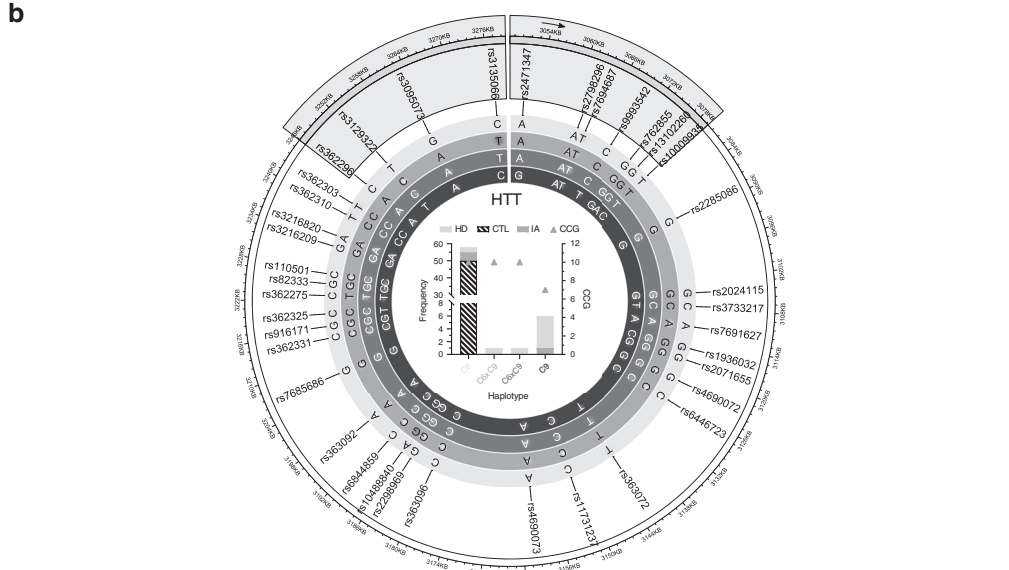
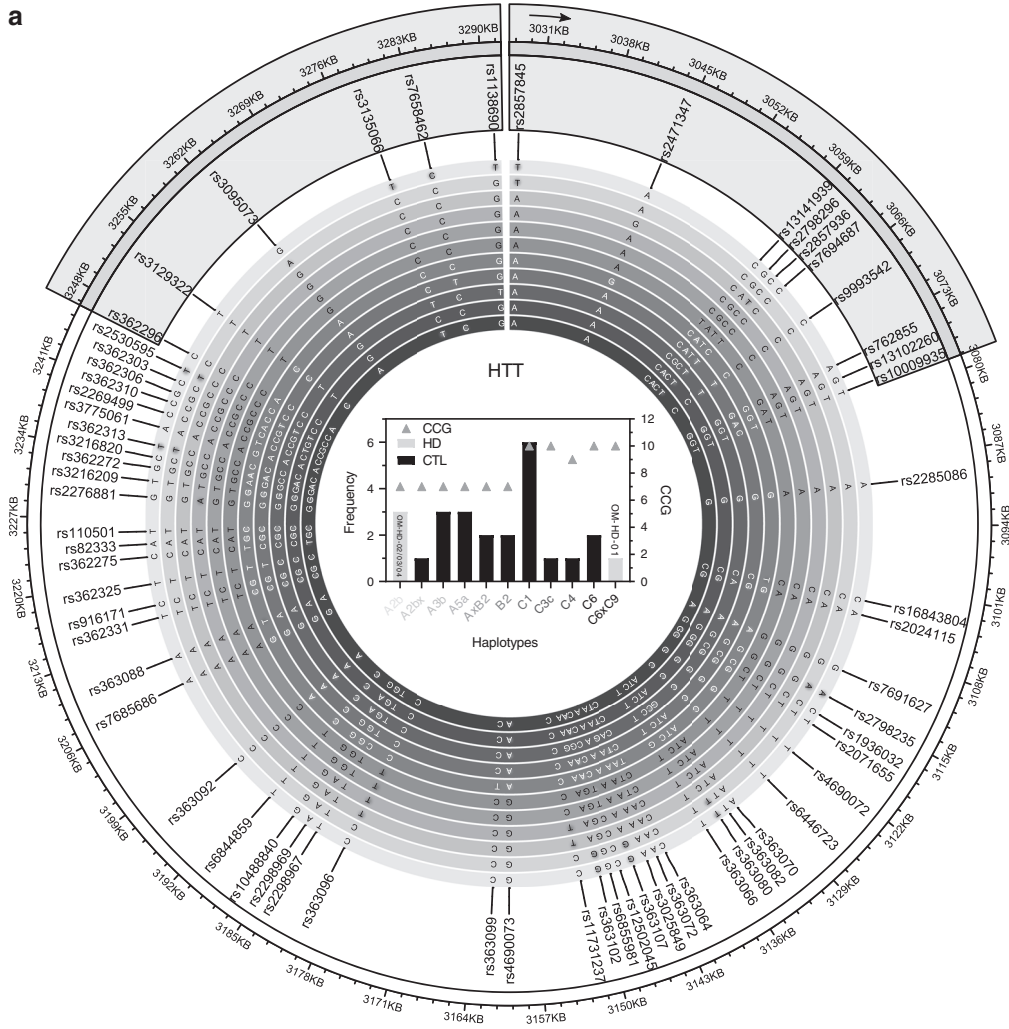
(48.8 vs 44.2, respectively;  $p < 0.0001$ ). Patients in the OM-HD-01 pedigree also had a significantly higher CAG age product (CAP)<sup>10</sup> score (532.3 vs. 415.4, respectively;  $p = 0.002$ ), a predictor of HD pathology severity (Table S3). The largest repeat number of ~200 CAG in one member of the OM-HD-01 pedigree was associated with an age at onset of 18 months. Onset of neurological symptoms was in the range expected for that given expanded allele size, according to Langbehn's model (Figure S3),<sup>11</sup> and results of the regression analysis revealed that CAG repeat length contributed to 94% of OM-HD-01 patients' age at onset ( $n = 25$ , mean age at onset = 33.1 years, SD = 13.6,  $R^2 = 0.94$ ;  $p < 0.0001$ ).

## DISCUSSION

Our study provides the largest HD cluster ever described in the Middle East, owing to its high number of intermarriages and subsequent increase in heritability risk. In our OM-HD-01 pedigree, we identified the oldest African family ancestor who transmitted a unique, as-yet unreported C6x9 HD haplotype that was of African, rather than European, origin, and that appeared to be associated with a large number of joHD cases.

The C6 haplotype appears to be rarely, if ever, associated with expanded HD alleles in Middle East and Caucasian European individuals.<sup>12</sup> However, in the South African Black population, C6 is frequently associated with expanded alleles in the intermediate to HD range (CAG >27 repeats).<sup>12</sup> As well, the C9 haplotype is a rare African-specific haplotype consistently observed only on HD and intermediate alleles,<sup>12</sup> providing further evidence for an African origin of this novel Omani HD allele. In contrast, we showed that most HD families from the Middle East, including three from Oman, shared an A2b haplotype that was similar to Europeans, consistent with findings from a recent comprehensive haplotype analysis of HD expanded alleles in populations from Southern Europe, South Asia, the Middle East, and admixed African ancestry.<sup>12</sup>

In our data set we only observed the C6x9 HD haplotype in another, unreported, patient of South African origin carrying 49 CAG repeats, an expanded repeat size length overlapping the high Omani CAG range (i.e., 48.8 expanded repeats on average). This observation perfectly fits with the data concerning the same haplotype we describe in OM-HD-01 patients, whose ancestor originated from a Southeast





**Fig. 2 Haplotypes on control and expanded *HTT* alleles in Omani and South African Black patients.** (a) Circular representation of 64 informative single-nucleotide polymorphisms (SNPs) spanning 263 Kb and overlapping with the *HTT* gene that were fully phased to CAG repeat length allowing identification of 4 unrelated HD chromosomes and 21 unrelated control chromosomes from Omani Huntington disease (HD) families. Three of four Omani HD alleles were found on the A2b *HTT* haplotype, previously shown to be the most common gene-spanning *HTT* haplotype in HD chromosomes of Middle Eastern ancestry. Each SNP is labeled with its rs number. Different alleles are indicated by different background colors. Two gray sectors mark SNPs that are present in the upstream region (right) and downstream region (left) of the *HTT* gene. A bar plot in the center represents the 11 haplotypes on the x-axis, with the same color codes of the surrounding lanes, the frequency of Omani controls and HD patients carrying the haplotypes on the left y-axis and the count of CCG repeat polymorphism for each haplotype on the right y-axis. (b) Circular representation of 38 informative SNPs, spanning 234 Kb and four haplotypes. Comparison illustrates the African origins of the rare C6x9 haplotype that refers to one southeast African black HD subject (green) and to OM-HD-01 patients (blue). The other three haplotypes refer to South African Black individuals. Informative SNPs with rs number and corresponding genotype in C6, C6x9 and C9 haplotypes: rs2471347: C6 (A), C6x9 (A), C9 (G); rs9993542: C6 (C), C6x9 (C), C9 (T); rs13102260: C6 (G), C6x9 (G), C9 (A); rs10009935: C6 (T), C6x9 (T), C9 (C); rs3733217: C6 (C), C6x9 (C), C9 (T); rs1936032: C6 (G), C6x9 (G), C9 (C); rs362325: C6 (C), C6x9 (C), C9 (T); rs362275: C6 (C), C6x9 (T), C9 (T); rs362310: C6 (T), C6x9 (C), C9 (C); rs362303: C6 (T), C6x9 (C), C9 (C); rs362296: C6 (C), C6x9 (A), C9 (A); rs3095073: C6 (G), C6x9 (A), C9 (A).

African tribe before moving to Ethiopia, and with the increased frequency of joHD patients in South Africa.<sup>13</sup> Whether this unique haplotype explains, at least in part or in association with other gene modifiers or CAG stretch modifications,<sup>14–16</sup> the occurrence of increase in the triplet instability and CAG repeat expansion, and aggregation of joHD cases within families of South Africa<sup>13</sup> or other areas of the world,<sup>17</sup> needs additional analyses and confirmation.

However, our study may represent an important starting point for further epidemiological analyses aiming to disclose the prevalence of HD and the frequency of joHD in African and Middle East regions.

Our findings confirm that, in human history, multiple separate and rare disease-causing *HTT* alleles have arisen spontaneously and independently in different geographical locations, and then have moved abroad via immigration. The question is whether this may affect HD frequency, HD severity (i.e., CAP score increase),<sup>8</sup> and clusters of atypical clinical presentations (i.e., joHD) worldwide. Findings from research into genes underlying other neurodegenerative disorders (e.g., C9orf72, which causes amyotrophic lateral sclerosis) have highlighted the occurrence of different predisposing haplotypes in different populations that contribute to disease prevalence and phenotype in diverse ways, while others (i.e., junkophilin-3 gene causing HD-like 2) support a common African origin for all patients.<sup>18,19</sup>

Of note, the existence of these different HD haplotypes in different populations of patients may have a significant impact on the development of novel targeted genetic therapies, such as the allele-specific antisense oligonucleotides, which target *HTT* messenger RNA of different haplotypes, suppressing the production of mutant or wild-type huntingtin protein. At present, allele-specific *HTT* gene-silencing techniques target only the most common, predominantly Caucasian, haplotypes.<sup>20</sup> Therefore, the occurrence of such rare haplotypes in isolated populations suggests nonselective therapies will likely represent an important future resource for some populations.

In conclusion, our evidence for a unique HD haplotype originating from Southeast Africa sheds light on an important missing piece of HD history, providing insights into how HD

spread from sub-Saharan Africa into the Middle East. Interestingly, the transmission seemed to occur during a similar historical timeframe as that reported for the spread of HD in Caucasian populations from Northern Europe to the Americas.

#### SUPPLEMENTARY INFORMATION

The online version of this article (<https://doi.org/10.1038/s41436-020-0895-1>) contains supplementary material, which is available to authorized users.

#### ACKNOWLEDGEMENTS

This study was financially supported by LIRH Foundation (funds from “5×1000” of taxes), by the Italian Ministry of Health (funds from Ricerca Finalizzata [RF-2016-02364123]), and by the Canadian Institutes of Health Research. We are grateful to Mohammed Al-Hajiri for careful and detailed historical assay of family OM-HD-01, and to Abigail Woollard, a professional medical writer funded by the LIRH Foundation, for providing editorial revisions to the manuscript.

#### DISCLOSURE

B.G.L. reports personal fees from AOP Orphan Pharmaceuticals, personal fees from Hoffmann-La Roche, grants from CHDI Foundation, other from Pfizer, other from Lundbeck, grants from JPND, grants from E-Rare, other from NeuraMetrix, personal fees from TEVA Pharmaceuticals, personal fees from Takeda, personal fees from Triplet TX, personal fees from PTC, personal fees from Sage Therapeutics, personal fees from Novartis, personal fees from Wave, outside the submitted work. F.S. provided consulting services and advisory board functions to Teva, Wave Pharma, Pfizer, Hoffmann-La Roche, Novartis, PTC Therapeutics, UCB, Oman Ministry of Health–Sultanate of Oman. He is cofounder, scientific officer, and consultant of the Italian League for Research on Huntington and related diseases (LIRH Foundation). The other authors declare no conflicts of interest.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES

1. Kay C, Collins JA, Wright GEB, et al. The molecular epidemiology of Huntington disease is related to intermediate allele frequency and haplotype in the general population. *Am J Med Genet B*. 2018;177:346–357.
2. Hayden MR, Hopkins HC, Macrea M, Beighton PH. The origin of Huntington's chorea in the Afrikaner population of South Africa. *S Afr Med J*. 1980;58:197–200.
3. Baine FK, Kay C, Ketelaar ME, et al. Huntington disease in the South African population occurs on diverse and ethnically distinct genetic haplotypes. *Eur J Hum Genet*. 2013;21:1120–1127.
4. Warby SC, Montpetit A, Hayden AR, et al. CAG expansion in the Huntington disease gene is associated with a specific and targetable predisposing haplogroup. *Am J Hum Genet*. 2009;84:351–366.
5. Warby SC, Visscher H, Collins JA, et al. HTT haplotypes contribute to differences in Huntington disease prevalence between Europe and East Asia. *Eur J Hum Genet*. 2011;19:561–566.
6. Li XY, Li HL, Dong Y, Gao B, et al. Haplotype analysis encompassing HTT gene in Chinese patients with Huntington's disease. *Eur J Neurol*. 2020;27:273–279.
7. Kremer B, Goldberg P, Andrew SE, et al. A worldwide study of the Huntington's disease mutation: the sensitivity and specificity of measuring CAG repeats. *N Engl J Med*. 1994;330:1401–1406.
8. Fusilli C, Migliore S, Mazza T, et al. Biological and clinical manifestations of juvenile Huntington's disease: a retrospective analysis. *Lancet Neurol*. 2018;17:986–993.
9. Kay C, Collins JA, Skotte NH, et al. Huntingtin haplotypes provide prioritized target panels for allele-specific silencing in Huntington disease patients of European ancestry. *Mol Ther*. 2015;23:1759–1771.
10. Penney JB Jr, Vonsattel JP, Macdonald ME, Gusella JF, Myers RH. CAG repeat number governs the development rate of pathology in Huntington's disease. *Ann Neurol*. 1997;41:689–692.
11. Langbehn DR, Brinkman RR, Falush D, Paulsen JS, Hayden MR, on behalf of an International Huntington's Disease Collaborative Group. A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length. *Clin Genet*. 2004;65:267–277.
12. Kay C, Collins JA, Caron NS, et al. A comprehensive haplotype targeting strategy for allele-specific HTT suppression in Huntington disease. *Am J Hum Genet*. 2019;105:1112–1125.
13. Hayden MR, MacGregor JM, Saffer DS, Beighton PH. The high frequency of juvenile Huntington's chorea in South Africa. *J Med Genet*. 1982;19:94–97.
14. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. CAG repeat not polyglutamine length determines timing of Huntington's disease onset. *Cell*. 2019;178:887–900.
15. Moss DJH, Pardiñas AF, Langbehn D, et al. Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. *Lancet Neurol*. 2017;16:701–711.
16. Ciosi M, Maxwell A, Cumming SA, et al. A genetic association study of glutamine-encoding DNA sequence structures, somatic CAG expansion, and DNA repair gene variants, with Huntington disease clinical outcomes. *EBioMedicine*. 2019;48:568–580.
17. Telenius H, Kremer HPH, Thellmann J, et al. Molecular analysis of juvenile Huntington disease: the major influence on (CAG) n repeat length is the sex of the affected parent. *Hum Mol Genet*. 1993;2:1535–1540.
18. Goldstein O, Gana-Weisz M, Nefussy B, et al. High frequency of C9orf72 hexanucleotide repeat expansion in amyotrophic lateral sclerosis patients from two founder populations sharing the same risk haplotype. *Neurobiol Aging*. 2018;64:e1–e7.
19. Krause A, Mitchell C, Essop F, et al. Junctophilin 3 (JPH3) expansion mutations causing Huntington disease like 2 (HDL2) are common in South African patients with African ancestry and a Huntington disease phenotype. *Am J Med Genet B*. 2015;168:573–585.
20. Tabrizi SJ, Leavitt BR, Landwehrmeyer GB, et al. Targeting Huntingtin expression in patients with Huntington's disease. *N Engl J Med*. 2019;380:2307–2316.