 doi:10.20944/preprints202211.0135.v1

Article

Rare Amyloid Precursor Protein Point Mutations Recapitulate Worldwide Migration and Admixture in Healthy Individuals: Implications For the Study of Neurodegeneration

Paolo Abondio [1,2], Francesco Bruno [3,4,*], Amalia Cecilia Bruni [3,°,+] and Donata Luiselli [1,°]

[1] Laboratory of Ancient DNA, Department of Cultural Heritage, University of Bologna, Via degli Ariani 1, 48121 Ravenna, Italy
[2] Laboratory of Molecular Anthropology and Center for Genome Biology, Department of Biological, Geological and Environmental Sciences, University of Bologna, Via Selmi 3, 40126 Bologna, Italy
[3] Regional Neurogenetic Center (CRN), Department of Primary Care, ASP Catanzaro, 88046 Lamezia Terme, Italy
[4] Association for Neurogenetic Research (ARN), 88046 Lamezia Terme, Italy
* Correspondence: francescobrunofb@gmail.com (F.B.)
° These authors share senior authorship
+ A.C.B. is former Director of the Regional Neurogenetic Center (CRN)

Abstract: Genetic discoveries related to Alzheimer's disease and other dementias have been performed using either large cohorts of affected subjects or multiple individuals from the same pedigree, therefore disregarding mutations in the context of healthy groups. Moreover, a large portion of studies so far have been performed on individuals of European ancestry, with a remarkable lack of epidemiological and genomic data from underrepresented populations. The present study aims at scanning 70 single-point mutations on the APP gene in a publicly available genetic dataset including 2.504 healthy individuals from 26 populations, and analyzing their distribution. Moreover, after gametic phase reconstruction, a pairwise comparison of the segments surrounding the mutations was performed to reveal patterns of haplotype sharing that could point to specific cross-population and cross-ancestry admixture events. Eight mutations have been detected in the worldwide dataset, with several of them being specific for a single individual, population or macroarea. Patterns of segment sharing reflect recent historical events of migration and admixture possibly linked to colonization campaigns. These observations reveal the population dynamics of the considered APP mutations in worldwide human groups, and support the development of ancestry-informed screening practices for the improvement of precision and personalized approaches to neurodegeneration and dementias.

Keywords: amyloid; Alzheimer's disease; neurodegenerative diseases; neurodegeneration; dementia; population genomics; migration; admixture; APP; neuropathology

1. Introduction

The Amyloid Precursor Protein (APP) is a type I integral membrane glycoprotein, formed by a large N-terminal glycosylated domain on the extracellular side and a smaller intracellular C-terminal domain. It is encoded by the *APP* gene, located on the long arm of chromosome 21 [1]. There are several isoforms of APP, but the predominant ones in the human brain, sorted by aminoacidic chain length, are APP695, APP751 and APP770 [2,3]. The physiological role of APP is not yet fully understood and remains open to debate in the scientific community, but several studies suggest that the protein, as expressed in the brain, seems important in regulating neuron development and survival, synaptic function and plasticity, cell adhesion and neuroprotection [3].

Under physiological conditions, APP is subject to two processing pathways, one non-amyloidogenic (α-secretase pathway) and one amyloidogenic (β-secretase pathway), the second of which leads to the production of damaging beta-amyloid (Aβ) peptides [4]. In

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 November 2022 doi:10.20944/preprints202211.0135.v1

the non-amyloidogenic pathway, the proteolytic cut of the α-secretase on the extracellular side of APP generates the soluble fragment sAPPα into the extracellular space. The remaining intracellular CTFα/CTF83 is further cleaved by the γ-secretase complex, leading to the formation of p3 and AICD fragments, that are further processed by caspases, producing Jcasp and C31 [5]. This pathway is non-amyloidogenic since α-secretase cuts within the sequence of the otherwise damaging Aβ peptide [6].

On the other hand, in the amyloidogenic pathway, the proteolytic cleavage of APP by β-secretase occurs at the level of the extreme N-terminal of the Aβ sequence, with the formation and subsequent release of the sAPPβ fragment into the extracellular domain, and of CTFβ/CTF99, which remains into the membrane. CTFβ is then cleaved by γ-secretase (a complex of four proteins, including presenilin, nicastrin, Aph-1 and Pen-2) which cuts in different positions at the C-terminal end of the Aβ sequence. The result of this processing is the formation of Aβ fragments between 39 and 43 amino acids in length - of which the most abundant are βA40 and βA42 - that are secreted, and of the previously mentioned AICD fragments, Jcasp and C31 [3,7,8].

To date, 52 pathogenic mutations in the *APP* gene can lead to Aβ deposition in the brain parenchyma and in cerebral blood vessels and at least twenty-six missense mutations, located within or immediately flanking the Aβ sequence, have been linked to Autosomal Alzheimer's Disease (ADAD) [9,10]. Interestingly, the rare coding variant APP A673T has been detected in Norwegian, Swedish, Icelandic, and Finnish populations and reported as associated with a reduced risk for AD [9]. The same mutation is absent in healthy elderly individuals of Asian descent and is much rarer in the U.S. population [11–13]. As far as we know, however, no more general population genetic studies have been performed on the *APP* gene in healthy individuals. Indeed, it must be noted that the majority of the crucial genetic discoveries related to AD and other dementias have been performed using either large cohorts of affected subjects or multiple individuals from the same pedigree, therefore disregarding their existence in the context of healthy groups. However, it is well known that isoforms of the same protein (e.g., APOE) have different functions based on the tissue in which are expressed (pleiotropic gene) and that the context conditions determine a variable expressivity of pathological condition, often linked to a protective function of these mutations in specific environments [14]. Moreover, a large portion of the studies so far have been performed on individuals of general European ancestry (and often excluding descendants of people from the Iberian Peninsula and Finland), with a remarkable lack of epidemiological and genomic data from underrepresented South American, African, South Asian and East Asian populations: this hinders the overall understanding of the impact of genetic mutations and their phenotypic manifestations in the context of neurodegeneration and dementia [15–18].

Given these premises, the present study aims at detecting and analyzing the distribution of 70 single-point mutations along the sequence of the *APP* gene in a publicly available dataset (identified as the "1KGP dataset") including the genetic sequences of 2.504 healthy individuals from 26 worldwide populations in 5 macroareas. Moreover, a pairwise comparison of the segments surrounding the mutations was performed to reveal patterns of haplotype sharing that could point to specific cross-population and cross-ancestry admixture events, therefore shedding light on the population dynamics of the considered *APP* mutations in worldwide human groups.

2. Results

2.1. Distribution of APP point mutations

When the selected 70 single nucleotide variants for the *APP* gene (see Supplementary Table S1) were searched along the sequence of chromosome 21 in the entire 1KGP dataset, a total of eight variants of interest were detected in 51 different individuals (30 females and 21 males; see Table 1 and Supplementary Tables S2 and S3). All mutation carriers (excluding the female TSI subject NA20259, who presents two copies of the KM670/671NL "Swedish" double mutation) are heterozygous, and two non-synonymous mutations

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 November 2022 doi:10.20944/preprints202211.0135.v1

(which change the identity of the aminoacidic residue in the protein sequence) are never carried together by the same individual. However, synonymous mutations G708G and I716I are both found together in eight individuals, representative of the African macroarea. Indeed, several mutations appear to be limited to a single individual (A713T in a male Mexican carrier; V604M in a female Japanese subject), a single population (E599K in three Finnish individuals) or a single macroarea (the aforementioned "Swedish" double mutation in seven people of European descent), while others seem to cross the boundaries of physical geography and population ancestry (c.*372A>G in three representatives of Europe, two of South Asia and one admixed American; G708G in eight African subjects, four Europeans and two people of Asian ancestry; S614G in 16 representatives of the African macroarea, but also in three admixed Americans). Given the overall rarity of these mutations and the fact that this study, focusing on healthy individuals at the time of data collection, finds a cumulative frequency for the mutations of interest at about 2%, there is the possibility that the supposedly deleterious alleles have not appeared by chance in multiple individuals of the same population. Indeed, investigating the sharing of identical segments around the mutation for the carriers of that specific allele could reveal, instead of random genetic variation, a common origin for each variant of interest.

Table 1. APP point mutations detected in the worldwide dataset.

Protein mutation	Location	Recorded codon change	dbSNP_ID	Position (GRCh37/hg19)
c.*372 A>G	3' UTR	Non-coding region	rs187940037	27253609
I716I[1]	Exon 17	ATC to ATA or ATT	rs145564988	27264097
A713T	Exon 17	GCG to ACG	rs63750066	27264108
G708G	Exon 17	GGC to GGT	rs148888161	27264121
KM670/671NL	Exon 16	AAG.ATG to AAT.CTG	rs281865161	27269938
S614G	Exon 14	AGC to GGC	rs112263157	27284122
V604M	Exon 14	GTG to ATG	rs199887707	27284152
E599K	Exon 14	GAA to AAA	rs140304729	27284167

[1] Mutation I716I was not present in the original 70 variants taken from the ALZforum database.

2.2. Segment sharing across individuals

For each of the eight detected mutations, segment sharing was investigated by comparing pairs of chromosome sequences among carriers and identifying a distribution of segment lengths that was then standardized as described in the following section 4.5. Statistically significant segments of unusually large size and carrying the mutation of interest have been detected for six of the APP aminoacidic variants (see Supplementary Tables S3 to S8), with the exclusion of mutations A713T and V604M, for which comparative analysis with other carriers could not be performed.

When comparing chromosomes carrying the genetic mutation causative of the E599K aminoacidic change, the three subjects belonging to the Finnish population all exhibit pairwise significant segment sharing (Supplementary Table S8). In particular, the two male individuals share a 227kb segment spanning almost 6.000 SNPs, while they both share the same 107kb region (including 2.543 point variants) with the representative Finnish female. All other comparisons (between a chromosome copy carrying the mutation and one not carrying it, or between two non-carrying chromosome copies) reveal much shorter segments (< 44kb) shared among pairs of individuals, and those that reach statistical significance are found either between the two male Finnish representatives, or the mutation-carrying and non-carrying chromosome copies of the female.

Comparing carriers for the mutation c.*372A>G, detected in the 3' untranslated region of the *APP* gene, it can be seen that significant segments are shared indiscriminately

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 November 2022 doi:10.20944/preprints202211.0135.v1

among pairs of individuals of European and South Asian ancestry, especially when comparing either two carrier chromosomes (length ranging from 26kb up to 647kb), or two non-carrier chromosomes (even though the latter comparisons provide much shorter significant segments, from 14.7kb up to 49kb). The admixed Colombian individual representative of American ancestry (CLM) shares a limited, but significant, carrier segment of 15.4kb with the Sri Lanka Tamil (STU) subject, and when comparing non-carrier and carrier chromosomes, an even shorter one (11kb) with both the aforementioned STU individual and a European carrier of Iberian ancestry (Supplementary Table S3).

Individuals with the synonymous protein sequence change G708G reveal that, when comparing chromosomes carrying the genetic mutation, significantly long segments are present either between subjects of African ancestry (including the admixed Caribbean individual, ACB) or people of European and South Asian descent, but no significant sharing is detected between African and non-African individuals, or with the Vietnamese carrier, KHV (see Supplementary Table S5). If chromosomes carrying the mutation are contrasted with chromosomes not carrying it, the non-carrying chromosome of the KHV subject does indeed share a significant DNA segment with the mutated chromosome of an Iberian individual, while the non-carrying chromosome of the ACB subject, as well as that of the Bengali individual representative of South Asia (BEB), show significant sharing with mutation-carrying chromosomes from the individuals of European descent (Supplementary Table S5).

The pattern of sharing for the mutation inducing the synonymous change I716I mirrors almost exactly that of G708G, as both are found specifically in the same individuals of African ancestry: in pairwise comparisons, the admixed individual ACB shares shorter genetic segments carrying the mutation than the other subjects (see Supplementary Tables S2 and S4). Indeed, seven of the individuals carry both mutations on the same chromosome copy, so that, the length of significant shared segments during pairwise comparison is exactly the same when considering the genetic surroundings of either mutation. However, the male individual belonging to the Luhya population from Kenya (LWK, of identifier NA19383) carries the mutations G708G and I716I on different copies of chromosome 21. He shares the smallest significant segments, when comparing sequences both carrying the I716I causative mutation, while the segments shared for the mutation G708G are not significant at all (Supplementary Table S4).

The genetic variation inducing the aminoacidic change S614G is carried predominantly by individuals of African ancestry. However, several significantly long segments carrying the mutation can be identified by comparing these sequences with those of the Peruvian (PEL) and, to a lesser degree, the Puerto Rican subject, both representatives of the American macroarea (see Supplementary Table S7). Similarly, when searching for shared segments between carrying and non-carrying chromosomes, the largest significantly long fragments are found between non-admixed people of African ancestry and the admixed representatives of Southwestern USA (ASW), the Caribbean (ACB), and the American macroarea (PEL, PUR and CLM). These segments, however, are all much shorter than the ones found when comparing two chromosomes carrying the mutation (see Supplementary Table S7).

The “Swedish” double mutation KM670/671NL, found only in individuals of European ancestry, reveals that the longest significant segment (>181 kb) between carrier chromosomes is shared by a representative of Central European ancestry (CEU) and a British/Scottish individual (GBR). Interestingly, the chromosome copies from the homozygous individual from Tuscany (TSI) share a much smaller segment between each other (16.8kb), and while one shares very large segments (126kb and 96kb) with non-carrying copies from a GBR and a CEU individual, respectively, the other only shares fragments equal or smaller than 16.8kb with any other non-carrying chromosome (see

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 November 2022 doi:10.20944/preprints202211.0135.v1

Supplementary Table S6). The following Figure 1 summarizes the putative origins and dynamics of the mutations described here.

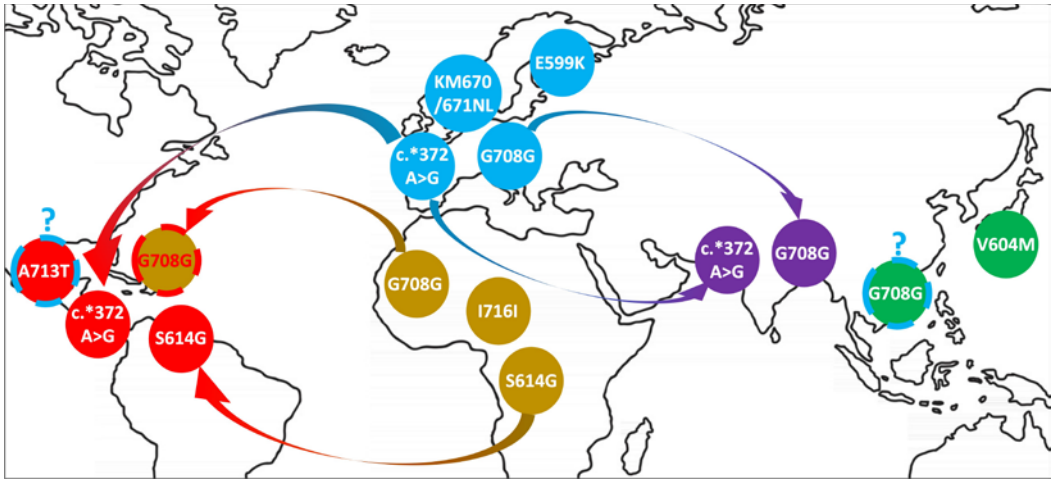

Figure 1. Possible origin and cross-population sharing of mutation-carrying segments. The Caribbean population of admixed origin is indicated by red text and red dashed borders. The putative European contributions to individuals carrying mutations A713T in America and G708G in Vietnam are indicated by a blue question mark and blue dashed borders.

3. Discussion

The present study was carried out to assess whether it was possible to detect genetic biomarkers of neurodegeneration (namely single nucleotide variants, putatively associated to AD, that would involve the sequence of the APP) in a dataset of healthy individuals, representatives of worldwide genetic diversity. Indeed, some of these mutations can be found, and the pattern of sharing for the surrounding genetic region can be analyzed through pairwise sequence comparison among carrier individuals, after gametic phase reconstruction and haplotype estimation are performed. Notably, each mutation detected in this study was uniquely shared by a group of individuals and, in general, two alleles of interest are never carried by the same subject (Supplementary Table S2). This suggests that these mutations, even in healthy individuals, do not tend to cluster or accumulate in the genetic sequence of *APP*, implying that the patterns of sharing for the different mutations can reveal separate instances of population dynamics. Taking advantage of notions from population genomics, it can be noted that several of the observed variants of interest present distributions that can be justified either by the random appearance of the same mutation in different populations and geographic areas, or by the movement of a genetic segment already carrying the mutation from a population to another through admixture, or sometimes by the appearance of a mutation after a non-carrying genetic segment has passed from a population to another. Moreover, it appears as though these patterns replicate instances of historical migration that took place over the last millennium.

For example, it is known that, although rare, the mutation A713T found here in an admixed Mexican subject has been already identified with different manifestations (early onset, late onset, familial, sporadic AD) in individuals or families from France [19,20], Spain [21], the United Kingdom [22] and Argentina [23]. Similarly, a major cluster has been found when analysing numerous unrelated families of Italian origin and ancestry (specifically from the Calabria region of Southern Italy [24]), where it has affected multiple generations with familial AD, often accompanied by cerebral amyloid angiopathy (CAA) and stroke as vascular manifestations of the disease [25–27]. Indeed, it has been also estimated that, given the size of the shared segments around it, this specific mutation may present a local Southern Italian origin from a common ancestor who lived there more than 1.000 years ago [28]. People of Iberian descent have been historically associated with the colonization of Central and South America after the "European discovery of the New

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 November 2022 doi:10.20944/preprints202211.0135.v1

World", during the Age of Discovery in the early 16th century [29]. It has not to be forgotten that sections of today's Italian and French territories were also part of the Spanish Empire, so that a non-negligible genetic input from non-Spanish European settlers of Mediterranean origin may have occurred at the time [30,31]. Similarly, it is estimated that during the so-called Italian diaspora (starting with the unification of Italy in 1861, with a third wave still occurring today) about 30 million people left the country, more than half of them before World War I (1914) [32–34]. The impact of this mass migration, which involved many Central and South American countries, can be observed through the largest communities of Italian descent existing today: about 47% of the people of Argentina (corresponding to roughly 19.7 million individuals), 35% of the people of Uruguay (1.2 millions) and 15% of current Brazilians (32 millions), not forgetting 17 million Americans of Italian ancestry also living in the United States [29,33]. So, alongside the plausible hypothesis of a random origin for the mutation A713T in the Mexican individual, there is the possibility that its admixed European ancestry may have provided the deleterious allele (Figure 1), although further studies are needed in order to prove it.

Mutation V604M, found here in a Japanese individual, has been recently identified in a total of three subjects from Thailand showing symptoms of early-onset AD [35–37]. Indeed, there is quite high genetic homogeneity among ethnic groups of Southeast Asian ancestry [38] so that a recent common ancestor for the mutation may exist even though the two areas are geographically distant (Figure 1). It is also possible that contacts leading to local admixture have taken place since the 13th century, first in the form of piracy and, three centuries later, through highly regulated commercial trade routes in Asian waters under the Ryukyu kingdom of Okinawa, which helped establish Japanese shipping colonies on foreign land facing the South China Sea and adjoining gulfs [39–41].

Synonymous variant G708G suggests that a more complex dynamics is involved. Indeed, segments detected among pairs of carrying chromosomes indicate that there is no sharing pattern across macroareas, suggesting that the same mutation may have appeared separately in Africa, Europe and East Asia at different times (Figure 1). This is not unexpected, as a neutral mutation that does not change the sequence, function or expression of a protein product may emerge multiple times in distinct populations, be preserved and pass along subsequent generations without being lost. However, the sharing of a very large significant segment surrounding the variant of interest between the non-carrying chromosome of the KHV subject and the mutated chromosome of the Iberian individual also aligns historically with the arrival of the first Portuguese missionaries in Vietnam and other Southeast Asian countries at the beginning of the 16th century [42,43], so that in this case a genetic segment may have been exchanged between the European settlers and the local communities around that time (Figure 1). A similar event may justify the sharing of a significantly long segment between the individual of Bengali ancestry and several European subjects, if the British rule over the Indian subcontinent is also factored in. In fact, the British East India company, created in 1600 to manage East-West trade across the Indian Ocean, had exerted commercial, military, and administrative power over Southeast Asia until the company's dissolution and the assumption of direct control by the British crown in 1858 [44]. Moreover, the admixed Caribbean subject, ACB, reveals a particularly intriguing story, as its non-carrying chromosome shows significant sharing with mutation-carrying chromosomes from the individuals of European descent, while the carrying chromosome shares a significant segment with other carrying chromosomes from African populations (Figure 1). Considering the aforementioned European colonial ventures of the 17th century (and, in particular, the French and British dominion over the "West Indies" after seizing them from Spain [45,46]), as well as the establishment of Jamaica as the center of the Anglo-American trade of African people across the Atlantic Ocean to use as slaves [47,48], it seems possible that at least one healthy Caribbean individual would show the effect of mixed ancestry in the different nature of her chromosome copies.

In this study we also report the presence of synonymous variant I716I only in individuals of African descent and, in seven out of eight cases, found on the exact same segment also carrying mutation G708G; only one male individual (NA19383) shows the two

 doi:10.20944/preprints202211.0135.v1

mutations on two different chromosome copies. To our knowledge, no specific study exists on this mutation that can help in the interpretation of the result. The gnomAD browser (https://gnomad.broadinstitute.org, last accessed 24 October 2022) indicates that the genetic change G>T causing the mutation is much more prevalent in people of African ancestry (186 out of 195 alleles detected over a sample size of 282.738 alleles), while the G>A genetic change is prevalent in non-Finnish European subjects (20 of 22 alleles over the previously mentioned sample size), which may explain why only African individuals with the mutation were found here. Moreover, the limited number of studies on the genetics of dementia and neurodegeneration in African populations impairs the possibility to analyze this result further, but at the same time provides a great opportunity for its future investigation [49]. We would like to suggest, however, that the existence of two synonymous mutations at close distance to one another may present an effect on the APP protein, even though its aminoacidic sequence is not changed. Several studies propose that changes in the nucleic acid triplets composing a codon (the sequence of three nitrous bases that underpins the incorporation of aminoacidic residues in a protein), may modulate specific properties by changing the speed of protein synthesis [50,51]. Indeed, codon usage bias (i.e., the difference in frequency occurrence of synonymous base triplets in coding DNA) implies in general that some codons may be preferentially used during protein synthesis as they allow for optimization of translation rates with high accuracy [50,52,53], and that this is a mechanism to balance mutation bias with natural selection, as seen in fast growing organisms with relatively small genomes [51,54,55]. The same phenomenon has been recently verified for human pluripotent embryonic stem cells, where codon bias is related to the different guanine-cytosine (GC) content of differentially expressed genes during stem cell differentiation [56]. In the specific case of G708G and I716I mutations, it would be interesting to experimentally test if codon bias has an influence on either translation speed or accuracy and, given the vectorial nature of *in vivo* protein synthesis, if this in turn impacts phenomena such as protein folding, membrane insertion and degradation.

4. Materials and Methods

4.1. Population data recovery

The distribution and variability of *APP* point mutations was analyzed in the publicly available Phase 3 dataset (hereby identified as "1KGP dataset") produced by the 1000 Genome Project Consortium [57], which includes 2.504 unrelated, self-reportedly healthy adult individuals (age > 18) representatives of 26 populations and classified in five macroareas (Africa, America, Europe, South Asia, East Asia). The dataset comprises populations collected in diaspora (where population ancestry and geographical location of the samples do not match) and groups of admixed ancestry, offering the opportunity to investigate the cross-population and cross-ancestry genetic context in which the mutations exist.

4.2. APP gene point mutations

Mutations for the *APP* gene studied in the context of neurodegeneration have been recovered from ALZforum (https://www.alzforum.org/mutations/app; last accessed 10 September 2022), an open-access resource for research news and information about Alzheimer's and related diseases, which catalogs scientific data about genetic variants and their influence on the manifestation and modulation of pathological phenotypes. Specifically, it manages a repository of genetic variants in genes linked to AD, with the goal of providing a list of variants that have been reported in the literature, ranging from causative to benign. For each variant, the salient clinical and neuropathologic features, as well as functional effects are reported. At last access, it includes the three genes (*APP, PSEN1, PSEN2*) associated with ADAD, plus two genes (*MAPT* and *TREM2*) with genetic associations to AD and related disorders. 70 single nucleotide variants, involved in both synonymous and non-synonymous aminoacidic changes of the canonical 770-residue long APP protein sequence, were recovered from the database (Supplementary Table S1);

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 8 November 2022 doi:10.20944/preprints202211.0135.v1

the mutations were checked for codon change sequence and NCBI dbSNP unique variant identifier (in the form rsX, where "X" is a unique numerical sequence; https://www.ncbi.nlm.nih.gov/snp/).

4.3. Dataset quality control

A standard quality control (QC) procedure was performed on the complete 1KGP dataset, which includes almost 80 million single nucleotide variants (as identified on the hg19 reference human sequence) covering all autosomal chromosomes (therefore, excluding the X and Y chromosomes, as well as mitochondrial DNA). This was performed before extracting variants limited to chromosome 21, where the gene of interest is located, so as not to introduce any bias. QC was performed with the PLINK version 1.9 software (https://www.cog-genomics.org/plink/) [58] and included the removal of variants or individuals with data missingness greater than 1% (--geno 0.01; --mind 0.01). A check for the respect of Hardy-Weinberg equilibrium for each variant was also performed, applying a Bonferroni correction for multiple testing to the standard threshold of 0.01, which is then divided by the number of variants (--hwe 1.8x10-10). Furthermore, ambiguous SNVs (carrying an A/T or C/G combination of alleles, for which the chromosome copy, and strand cannot be univocally defined) were removed and all remaining variants on chromosome 21 were extracted (--chr 21) for haplotype reconstruction.

4.4. Data phasing and haplotype estimation

In order to assess the presence of the APP mutation in the context of the surrounding variants, to detect finer-scale relationship patterns and to compare chromosomal sequences for further statistical analysis (see the following paragraph 4.5), a data phasing procedure and haplotype estimation was performed in order to define which allele of every variant is located on which copy of the chromosome, for each of the 2.504 individuals of the 1KGP dataset. To do so, information about the ancestral or derived nature of each variant was deduced by using a reconstructed reference human genome sequence as a guide for distinguishing between ancestral and derived alleles. In particular, the ancestral/derived state of each allele in such a reference sequence was previously assigned by aligning it with the Ensembl Compara 6 primates EPO genome sequences [59], and only alleles present in all the compared genomes were considered as ancestral. Haplotype estimation is finally performed using the SHAPEIT software version 1.9 (https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit) [60,61] with default parameters settings and the HapMap phase 3 recombination map for chromosome 21. Reconstruction was performed on a total of 870.655 SNPs.

4.5. Segment detection and statistical validation

Detection of shared chromosomal segments harboring the identified mutations in the 1KGP chromosome 21 dataset was performed separately for each single mutation and limited to the individuals who carry it. Pairs of chromosomes (both with the mutation, or one with it and one without it) were compared point-by-point along their entire length through an ad hoc Python script (version 2.7.13; https://www.python.org), and a "segment" is considered as each continuous sequence of at least two alleles that are identical in the compared chromosome sequences. This generates a distribution of segment lengths, where "length" is defined as the absolute value of the difference in physical positions between the extremities of each segment. To statistically define if the shared segment carrying the mutation is significantly longer than the others (and, as such, can be considered uncommonly shared between individuals), the distribution of lengths obtained for each pairwise chromosome comparison is standardized by subtracting to each length the average value of the distribution, and dividing the result by the standard deviation of the same distribution. The new obtained distribution has the characteristics of a standardized distribution centered around the zero value, where the normalized value of each length tells how many standard deviations each segment falls from the

 doi:10.20944/preprints202211.0135.v1

average of the distribution. Absolute normalized values above two standard deviations suggest that the associated segment falls among the extreme values of the distribution, which indicates that the sharing of segments between two individuals, with respect to all the other shared segments along the chromosome, can be considered significant.

5. Conclusions

The present study, involving the detection and distribution analysis of rare *APP* point mutations in a publicly available dataset including individuals from 26 global populations, representatives of human genetic variability, reveals that variants usually analysed in the context of their pathological manifestations may indeed be recovered from large datasets of healthy subjects. Furthermore, significant sharing of chromosomal segments (haplotypes) surrounding the variants of interest reveals cross-population contact and peculiar cross-ancestry dynamics that can be at least partially brought back to worldwide phenomena of migration, colonization and admixture in recent history. Indeed, individual and population ancestry, driven by past and recent microevolutionary dynamics, are relevant for the presence of specific causative mutations in several human groups, so that they cannot be discarded as factors determining the evaluation and contextualization of genetic variants associated with neurodegenerative diseases. Moreover, even if the individual is carrying a causative mutation, its penetrance and expression are modulated both by its genetic background and the environmental context, so that the same mutation in different populations may not have the same phenotypic manifestation. Given the rate of both historical and contemporary worldwide migrations, this knowledge may be useful for the development of ancestry-informed screening practices and the improvement of precision and personalized approaches to neurodegeneration and dementias.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Table S1: APP point mutations recovered from the ALZforum database; Table S2: APP point mutations recovered in the 1KGP dataset; Table S3: Pairwise segment sharing for mutation c.*372A>G; Table S4: Pairwise segment sharing for mutation I716I; Table S5: Pairwise segment sharing for mutation G708G; Table S6: Pairwise segment sharing for the "Swedish" mutation KL670/671MN; Table S7: Pairwise segment sharing for mutation S614G; Table S8: Pairwise segment sharing for mutation E599K.

Author Contributions: Conceptualization, P.A. and D.L.; methodology, P.A.; formal analysis, P.A. and F.B.; investigation, F.B. and A.C.B.; resources, D.L. and A.C.B.; data curation, P.A. and F.B.; writing—original draft preparation, P.A. and F.B.; writing—review and editing, A.C.B. and D.L.; visualization, P.A.; supervision, A.C.B. and D.L.; project administration, D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable: the present study only includes freely available, public, and anonymized data.

Informed Consent Statement: Not applicable: the present study only includes freely available, public, and anonymized data.

Data Availability Statement: All data used in this study are in the public domain and their sources are indicated in the main text; no new data has been produced for this study.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Kirkitadze, M.D.; Kowalska, A. Molecular Mechanisms Initiating Amyloid Beta-Fibril Formation in Alzheimer's Disease. *Acta Biochim Pol* **2005**, *52*, 417–423.
2. Mattson, M.P. Pathways towards and Away from Alzheimer's Disease. *Nature* **2004**, *430*, 631–639, doi:10.1038/nature02621.

3. Müller, U.C.; Deller, T.; Korte, M. Not Just Amyloid: Physiological Functions of the Amyloid Precursor Protein Family. *Nat Rev Neurosci* **2017**, *18*, 281–298, doi:10.1038/nrn.2017.29.
4. Kojro, E.; Fahrenholz, F. The Non-Amyloidogenic Pathway: Structure and Function of Alpha-Secretases. *Subcell Biochem* **2005**, *38*, 105–127, doi:10.1007/0-387-23226-5_5.
5. Lopez Sanchez, M.I.G.; van Wijngaarden, P.; Trounce, I.A. Amyloid Precursor Protein-Mediated Mitochondrial Regulation and Alzheimer's Disease. *Br J Pharmacol* **2019**, *176*, 3464–3474, doi:10.1111/bph.14554.
6. Bandyopadhyay, S.; Hartley, D.M.; Cahill, C.M.; Lahiri, D.K.; Chattopadhyay, N.; Rogers, J.T. Interleukin-1alpha Stimulates Non-Amyloidogenic Pathway by Alpha-Secretase (ADAM-10 and ADAM-17) Cleavage of APP in Human Astrocytic Cells Involving P38 MAP Kinase. *J Neurosci Res* **2006**, *84*, 106–118, doi:10.1002/jnr.20864.
7. Nalivaeva, N.N.; Turner, A.J. Targeting Amyloid Clearance in Alzheimer's Disease as a Therapeutic Strategy. *Br J Pharmacol* **2019**, *176*, 3447–3463, doi:10.1111/bph.14593.
8. Hampel, H.; Hardy, J.; Blennow, K.; Chen, C.; Perry, G.; Kim, S.H.; Villemagne, V.L.; Aisen, P.; Vendruscolo, M.; Iwatsubo, T.; et al. The Amyloid-β Pathway in Alzheimer's Disease. *Mol Psychiatry* **2021**, *26*, 5481–5503, doi:10.1038/s41380-021-01249-0.
9. Tcw, J.; Goate, A.M. Genetics of β-Amyloid Precursor Protein in Alzheimer's Disease. *Cold Spring Harb Perspect Med* **2017**, *7*, a024539, doi:10.1101/cshperspect.a024539.
10. Breijyeh, Z.; Karaman, R. Comprehensive Review on Alzheimer's Disease: Causes and Treatment. *Molecules* **2020**, *25*, 5789, doi:10.3390/molecules25245789.
11. Jonsson, T.; Atwal, J.K.; Steinberg, S.; Snaedal, J.; Jonsson, P.V.; Bjornsson, S.; Stefansson, H.; Sulem, P.; Gudbjartsson, D.; Maloney, J.; et al. A Mutation in APP Protects against Alzheimer's Disease and Age-Related Cognitive Decline. *Nature* **2012**, *488*, 96–99, doi:10.1038/nature11283.
12. Kero, M.; Paetau, A.; Polvikoski, T.; Tanskanen, M.; Sulkava, R.; Jansson, L.; Myllykangas, L.; Tienari, P.J. Amyloid Precursor Protein (APP) A673T Mutation in the Elderly Finnish Population. *Neurobiol Aging* **2013**, *34*, 1518.e1-3, doi:10.1016/j.neurobiolaging.2012.09.017.
13. Liu, Y.-W.; He, Y.-H.; Zhang, Y.-X.; Cai, W.-W.; Yang, L.-Q.; Xu, L.-Y.; Kong, Q.-P. Absence of A673T Variant in APP Gene Indicates an Alternative Protective Mechanism Contributing to Longevity in Chinese Individuals. *Neurobiol Aging* **2014**, *35*, 935.e11-12, doi:10.1016/j.neurobiolaging.2013.09.023.
14. Abondio, P.; Sazzini, M.; Garagnani, P.; Boattini, A.; Monti, D.; Franceschi, C.; Luiselli, D.; Giuliani, C. The Genetic Variability of APOE in Different Human Populations and Its Implications for Longevity. *Genes (Basel)* **2019**, *10*, E222, doi:10.3390/genes10030222.
15. Adams, H.H.H.; Evans, T.E.; Terzikhan, N. The Uncovering Neurodegenerative Insights Through Ethnic Diversity Consortium. *The Lancet Neurology* **2019**, *18*, 915, doi:10.1016/S1474-4422(19)30324-2.
16. Mills, M.C.; Rahal, C. A Scientometric Review of Genome-Wide Association Studies. *Commun Biol* **2019**, *2*, 9, doi:10.1038/s42003-018-0261-x.
17. Sirugo, G.; Williams, S.M.; Tishkoff, S.A. The Missing Diversity in Human Genetic Studies. *Cell* **2019**, *177*, 26–31, doi:10.1016/j.cell.2019.02.048.
18. Raman, R.; Quiroz, Y.T.; Langford, O.; Choi, J.; Ritchie, M.; Baumgartner, M.; Rentz, D.; Aggarwal, N.T.; Aisen, P.; Sperling, R.; et al. Disparities by Race and Ethnicity Among Adults Recruited for a Preclinical Alzheimer Disease Trial. *JAMA Netw Open* **2021**, *4*, e2114364, doi:10.1001/jamanetworkopen.2021.14364.
19. Carter, D.A.; Desmarais, E.; Bellis, M.; Campion, D.; Clerget-Darpoux, F.; Brice, A.; Agid, Y.; Jaillard-Serradt, A.; Mallet, J. More Missense in Amyloid Gene. *Nat Genet* **1992**, *2*, 255–256, doi:10.1038/ng1292-255.
20. Lanoiselée, H.-M.; Nicolas, G.; Wallon, D.; Rovelet-Lecrux, A.; Lacour, M.; Rousseau, S.; Richard, A.-C.; Pasquier, F.; Rollin-Sillaire, A.; Martinaud, O.; et al. APP, PSEN1, and PSEN2 Mutations in Early-Onset Alzheimer Disease: A Genetic Screening Study of Familial and Sporadic Cases. *PLoS Med* **2017**, *14*, e1002270, doi:10.1371/journal.pmed.1002270.
21. Armstrong, J.; Boada, M.; Rey, M.J.; Vidal, N.; Ferrer, I. Familial Alzheimer Disease Associated with A713T Mutation in APP. *Neuroscience Letters* **2004**, *370*, 241–243, doi:10.1016/j.neulet.2004.08.026.
22. Barber, I.S.; García-Cárdenas, J.M.; Sakdapanichkul, C.; Deacon, C.; Zapata Erazo, G.; Guerreiro, R.; Bras, J.; Hernandez, D.; Singleton, A.; Guetta-Baranes, T.; et al. Screening Exons 16 and 17 of the Amyloid Precursor Protein Gene in Sporadic Early-Onset Alzheimer's Disease. *Neurobiology of Aging* **2016**, *39*, 220.e1-220.e7, doi:10.1016/j.neurobiolaging.2015.12.011.
23. Fernández Suarez, M.C.; Brusco, I.; Damasso, C.; Olivar, N.; Morelli, L.; Russo, G. Heterozygous App A713t Mutation Carrier with Inflammatoy Amyloid Angiopathy and Family History of Alzheimer´s Disease. First Case in Argentina. *JNSK* **2019**, *9*, doi:10.15406/jnsk.2019.09.00353.
24. Bruno, F.; Laganà, V.; Di Lorenzo, R.; Bruni, A.C.; Maletta, R. Calabria as a Genetic Isolate: A Model for the Study of Neurodegenerative Diseases. *Biomedicines* **2022**, *10*, 2288, doi:10.3390/biomedicines10092288.
25. Rossi, G.; Giaccone, G.; Maletta, R.; Morbin, M.; Capobianco, R.; Mangieri, M.; Giovagnoli, A.R.; Bizzi, A.; Tomaino, C.; Perri, M.; et al. A Family with Alzheimer Disease and Strokes Associated with A713T Mutation of the APP Gene. *Neurology* **2004**, *63*, 910–912, doi:10.1212/01.WNL.0000137048.80666.86.
26. Bernardi, L.; Geracitano, S.; Colao, R.; Puccio, G.; Gallo, M.; Anfossi, M.; Frangipane, F.; Curcio, S.A.M.; Mirabelli, M.; Tomaino, C.; et al. AβPP A713T Mutation in Late Onset Alzheimer's Disease with Cerebrovascular Lesions. *JAD* **2009**, *17*, 383–389, doi:10.3233/JAD-2009-1061.

27. Conidi, M.E.; Bernardi, L.; Puccio, G.; Smirne, N.; Muraca, M.G.; Curcio, S.A.M.; Colao, R.; Piscopo, P.; Gallo, M.; Anfossi, M.; et al. Homozygous Carriers of APP A713T Mutation in an Autosomal Dominant Alzheimer Disease Family. *Neurology* **2015**, *84*, 2266–2273, doi:10.1212/WNL.0000000000001648.
28. Abondio, P.; Sarno, S.; Giuliani, C.; Laganà, V.; Maletta, R.; Bernardi, L.; Bruno, F.; Colao, R.; Puccio, G.; Frangipane, F.; et al. Amyloid Precursor Protein A713T Mutation in Calabrian Patients with Alzheimer's Disease: A Population Genomics Approach to Estimate Inheritance from a Common Ancestor. *Biomedicines* **2021**, *10*, 20, doi:10.3390/biomedicines10010020.
29. Smith, P.H.; Green, J.N. *Modern Latin America*; Ninth edition.; Oxford University Press: New York, 2019; ISBN 978-0-19-067465-6.
30. *Global Goods and the Spanish Empire, 1492 - 1824: Circulation, Resistance and Diversity*; Aram, B., Yun Casalilla, B., Eds.; Palgrave Macmillan: Basingstoke, Hampshire, 2014; ISBN 978-1-137-32405-4.
31. *The Spanish Empire: A Historical Encyclopedia*; Tarver Denova, H.M., Slape, E., Eds.; Empires of the world; ABC-CLIO: Santa Barbara, California, 2016; ISBN 978-1-61069-421-6.
32. *The Italian Diaspora: Migration across the Globe*; Pozzetta, G.E., Ed.; Studies in Ethnic and immigration history; Multicultural History Society: Ontario, 1992; ISBN 978-0-919045-59-0.
33. *Italian Mobilities*; Ben-Ghiat, R., Hom, S.M., Eds.; 1 Edition.; Routledge, Taylor & Francis Group: London ; New York, 2015; ISBN 978-1-138-77814-6.
34. Galante, J.S. *On the Other Shore: The Atlantic Worlds of Italians in South America during the Great War*; University of Nebraska Press: Lincoln, 2022; ISBN 978-1-4962-0791-3.
35. Giau, V.V.; Senanarong, V.; Bagyinszky, E.; Limwongse, C.; An, S.S.A.; Kim, S. Identification of a Novel Mutation in APP Gene in a Thai Subject with Early-Onset Alzheimer's Disease. *NDT* **2018**, *Volume 14*, 3015–3023, doi:10.2147/NDT.S180174.
36. Giau, V.; Senanarong, V.; Bagyinszky, E.; An, S.; Kim, S. Analysis of 50 Neurodegenerative Genes in Clinically Diagnosed Early-Onset Alzheimer's Disease. *IJMS* **2019**, *20*, 1514, doi:10.3390/ijms20061514.
37. Giau, V.V.; Bagyinszky, E.; Youn, Y.C.; An, S.S.A.; Kim, S. APP, PSEN1, and PSEN2 Mutations in Asian Patients with Early-Onset Alzheimer Disease. *IJMS* **2019**, *20*, 4757, doi:10.3390/ijms20194757.
38. Tagore, D.; Aghakhanian, F.; Naidu, R.; Phipps, M.E.; Basu, A. Insights into the Demographic History of Asia from Common Ancestry and Admixture in the Genomic Landscape of Present-Day Austroasiatic Speakers. *BMC Biol* **2021**, *19*, 61, doi:10.1186/s12915-021-00981-x.
39. *The East Asian Maritime World 1400-1800: Its Fabrics of Power and Dynamics of Exchanges*; Schottenhammer, A., Ed.; East Asian economic and socio-cultural studies. East Asian maritime history; Harrassowitz: Wiesbaden, 2007; ISBN 978-3-447-05474-4.
40. Akamine, M. *The Ryukyu Kingdom: Cornerstone of East Asia*; Huey, R.N., Ed.; Paperback edition.; University of Hawai'i Press: Honolulu, 2019; ISBN 978-0-8248-5517-8.
41. *East Asia in the World: Twelve Events That Shaped the Modern International Order*; Haggard, S., Kang, D.C., Eds.; Cambridge University Press: Cambridge, United Kingdom ; New York, NY, 2020; ISBN 978-1-108-80740-1.
42. Winius, G.D. *Studies on Portuguese Asia, 1495-1689*; Variorum collected studies series; Ashgate: Aldershot ; Burlington, USA, 2001; ISBN 978-0-86078-872-0.
43. Subrahmanyam, S. *The Portuguese Empire in Asia, 1500-1700: A Political and Economic History*; 2nd ed.; Wiley: Hoboken, NJ, 2012; ISBN 978-1-118-27402-6.
44. Lawson, P. *The East India Company: A History*; Studies in modern history; Longman: London ; New York, 1993; ISBN 978-0-582-07386-9.
45. PITMAN, F.W. *Development of the British West Indies: 1700-1763*; ROUTLEDGE: Place of publication not identified, 2020; ISBN 978-0-367-14259-9.
46. Collins, J.K. *Tracing British West Indian Slavery Laws: A Comparative Analysis of Legal Transplants*; Routledge studies in comparative legal history; Routledge: Abingdon, Oxon [UK] ; New York, NY, 2021; ISBN 978-1-00-051567-1.
47. Delle, J.A. *The Colonial Caribbean: Landscapes of Power in Jamaica's Plantation System*; Cambridge University Press: New York, 2014; ISBN 978-1-139-02404-4.
48. *The Colonial Landscape of the British Caribbean*; Leech, R., Leech, P., Eds.; The Society for Post-Medieval Archaeology monograph; The Boydell Press: Woodbridge, 2021; ISBN 978-1-78327-565-6.
49. Akinyemi, R.O.; Yaria, J.; Ojagbemi, A.; Guerchet, M.; Okubadejo, N.; Njamnshi, A.K.; Sarfo, F.S.; Akpalu, A.; Ogbole, G.; Ayantayo, T.; et al. Dementia in Africa: Current Evidence, Knowledge Gaps, and Future Directions. *Alzheimer's & Dementia* **2022**, *18*, 790–809, doi:10.1002/alz.12432.
50. Yu, C.-H.; Dang, Y.; Zhou, Z.; Wu, C.; Zhao, F.; Sachs, M.S.; Liu, Y. Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-Translational Protein Folding. *Molecular Cell* **2015**, *59*, 744–754, doi:10.1016/j.molcel.2015.07.018.
51. Qian, W.; Yang, J.-R.; Pearson, N.M.; Maclean, C.; Zhang, J. Balanced Codon Usage Optimizes Eukaryotic Translational Efficiency. *PLoS Genet* **2012**, *8*, e1002603, doi:10.1371/journal.pgen.1002603.
52. Hanson, G.; Coller, J. Codon Optimality, Bias and Usage in Translation and MRNA Decay. *Nat Rev Mol Cell Biol* **2018**, *19*, 20–30, doi:10.1038/nrm.2017.91.
53. Liu, Y. A Code within the Genetic Code: Codon Usage Regulates Co-Translational Protein Folding. *Cell Commun Signal* **2020**, *18*, 145, doi:10.1186/s12964-020-00642-6.
54. Arella, D.; Dilucca, M.; Giansanti, A. Codon Usage Bias and Environmental Adaptation in Microbial Organisms. *Mol Genet Genomics* **2021**, *296*, 751–762, doi:10.1007/s00438-021-01771-4.

55. Sharp, P.M.; Emery, L.R.; Zeng, K. Forces That Influence the Evolution of Codon Bias. *Phil. Trans. R. Soc. B* **2010**, *365*, 1203–1212, doi:10.1098/rstb.2009.0305.
56. Bornelöv, S.; Selmi, T.; Flad, S.; Dietmann, S.; Frye, M. Codon Usage Optimization in Pluripotent Embryonic Stem Cells. *Genome Biol* **2019**, *20*, 119, doi:10.1186/s13059-019-1726-z.
57. The 1000 Genomes Project Consortium; Corresponding authors; Auton, A.; Abecasis, G.R.; Steering committee; Altshuler, D.M.; Durbin, R.M.; Abecasis, G.R.; Bentley, D.R.; Chakravarti, A.; et al. A Global Reference for Human Genetic Variation. *Nature* **2015**, *526*, 68–74, doi:10.1038/nature15393.
58. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.R.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.W.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **2007**, *81*, 559–575, doi:10.1086/519795.
59. Herrero, J.; Muffato, M.; Beal, K.; Fitzgerald, S.; Gordon, L.; Pignatelli, M.; Vilella, A.J.; Searle, S.M.J.; Amode, R.; Brent, S.; et al. Ensembl Comparative Genomics Resources. *Database* **2016**, *2016*, bav096, doi:10.1093/database/bav096.
60. Delaneau, O.; Marchini, J.; Zagury, J.-F. A Linear Complexity Phasing Method for Thousands of Genomes. *Nat Methods* **2012**, *9*, 179–181, doi:10.1038/nmeth.1785.
61. Delaneau, O.; Zagury, J.-F.; Marchini, J. Improved Whole-Chromosome Phasing for Disease and Population Genetic Studies. *Nat Methods* **2013**, *10*, 5–6, doi:10.1038/nmeth.2307.