

Validation of Suspected Somatic Single Nucleotide Variations in the Brain of Alzheimer's Disease Patients

Alberto Gomez-Ramos^{a,b}, Angel J. Picher^c, Esther García^{a,b}, Patricia Garrido^c, Felix Hernandez^{a,b}, Eduardo Soriano^{a,d,e,f,*} and Jesús Avila^{a,b,*}

^a*Centro de Investigación Biomédica en Red de Enfermedades Neurodegenerativas (CIBERNED), ISCIII, Madrid, Spain*

^b*Centro de Biología Molecular Severo Ochoa (CSIC-UAM), Madrid, Spain*

^c*Sygnis S.L.U. Parque Científico de Madrid. Cantoblanco, Madrid, Spain*

^d*Department of Cell Biology and Institute of Neurosciences, University of Barcelona, Barcelona, Spain*

^e*Vall d'Hebrón Institut de Recerca (VHIR), Barcelona, Spain*

^f*ICREA Academia, Barcelona, Spain*

Accepted 25 November 2016

Abstract. Next-generation sequencing techniques and genome-wide association study analyses have provided a huge amount of data, thereby enabling the identification of DNA variations and mutations related to disease pathogenesis. New techniques and software tools have been developed to improve the accuracy and reliability of this identification. Most of these tools have been designed to discover and validate single nucleotide variants (SNVs). However, in addition to germ-line mutations, human tissues bear genomic mosaicism, which implies that somatic events are present only in low percentages of cells within a given tissue, thereby hindering the validation of these variations using standard genetic tools. Here we propose a new method to validate some of these somatic mutations. We combine a recently developed software with a method that cuts DNA by using restriction enzymes at the sites of the variation. The non-cleaved molecules, which bear the SNV, can then be amplified and sequenced using Sanger's technique. This procedure, which allows the detection of alternative alleles present in as few as 10% of cells, could be of value for the identification and validation of low frequency somatic events in a variety of tissues and diseases.

Keywords: Alzheimer's disease, identification of nucleotide variations, Sanger sequencing, somatic mutations, Virimid software

INTRODUCTION

A small number of Alzheimer's disease (AD) cases are caused by an inherited mutation in one of three genes (APP, PSEN1, and PSEN2) [1]. However, the vast majority of patients have the sporadic form of the disease, which is related to the presence of

several genetic and non-genetic risk factors [2]. Since somatic mutations in single human neurons have been identified [3, 4], it has been proposed that somatic gene variations present in brain cells facilitate the development of the disease [5, 6]. These variations can arise during development or aging, as in the case of other disorders such as cancer, and they may represent mosaic genomic heterogeneity [7].

It is difficult to confirm the detection of somatic gene variations by standard procedures like Sanger's method when the number of cells bearing the particular gene variation or mutation is very low compared with the rest of the cells. Indeed, Sanger sequencing

*Correspondence to: Jesús Avila, Centro de Biología Molecular Severo Ochoa (CSIC-UAM), 28049 Madrid, Spain. Tel.: +34 911964564; Fax: +34 911964420; E-mail: javila@cbm.csic.es and Eduardo Soriano, Department of Cell Biology, University of Barcelona, 08028 Barcelona, Spain. Tel.: +34 934037117; E-mail: esoriano@ub.edu.

allows the detection of somatic gene variations with an allelic frequency of 20% or higher [8] in a given sample. In this regard, other techniques, such as Illumina sequencing, have been tested [6]. However, this method can introduce errors in reading sequencing alignments that can interfere with the identification of true somatic variations [9, 10]. In an attempt to decrease these errors, several approaches have emerged [5, 10, 11]. However, it continues to be difficult to apply Sanger's sequencing for validation purposes when variations are present in only a few cells or when comparing samples from two different tissues from the same individual and only one bears the somatic gene variation.

Here we first developed a new method involving Illumina sequencing and validation that allows the characterization of somatic variations present in a few brain cells and absent in other tissues like blood, of a single subject. It is difficult to accurately identify a somatic SNV in the exome of a specific cell in an environment containing many cells lacking that particular SNV. New methods have recently been proposed [5, 10–14], and new software, like Virmid [15], has been developed to favor the detection of somatic mutations without impurity interferences and to avoid errors caused by non-suitable read alignments. In this way, the main aim of this work is the development of a new method for the analysis of somatic variants.

In the present study, we describe a method to identify exomic brain-specific somatic SNVs by Illumina sequencing. The method is based on the following: a) use of Virmid software [15] for data processing; b) removal of the DNA lacking the SNV of interest by digestion with a specific restriction nuclease; and c) amplification and validation of the DNA containing the SNV by Sanger sequencing. This amplification is achieved by treating the DNA samples with specific restriction endonucleases that recognize only those sequences in which the SNV of interest is absent.

These samples are then cleaved. Finally, the DNA bearing the SNV is amplified and subjected to Sanger sequencing.

A limitation of this procedure is that the restriction nuclease used to eliminate bulk DNA may not always be available; however, when we performed the nuclease step, there was good correlation between the bioinformatics processing and the data obtained by Sanger sequencing after nuclease treatment. This finding indicates that this method is suitable to identify brain SNVs that occur in neurodegenerative disorders like AD, and, in general terms, to identify the SNVs present at low proportions in a particular tissue.

MATERIALS AND METHODS

Samples from donors were obtained from the Spanish Brain Bank (*Banco de Tejidos CIEN* [BT-CIEN], <http://bt.fundacioncien.es/>) and from the *Biobanco del Sistema Sanitario Público de Andalucía* (<http://www.juntadeandalucia.es/salud/biobanco/>). When alive, donors gave their written informed consent. The tissues were obtained using protocols approved by the ethical committee of the two aforementioned organizations. Our work was previously approved by the ethical committee of our center (*Comité de Ética de la Investigación conjunto CNB-CBMSO*, <http://www.cnb.csic.es/cei/>). The methods were carried out in accordance with the approved guidelines. DNA was extracted using Qiagen kits and following the manufacturer's instructions.

The list of donors is indicated in Supplementary Table 1.

Tissue sample preparation

Hippocampal tissues from donors were obtained as follows: postmortem tissues were obtained through a

Table 1
SNVs from AD patients validated by the present method

Sample	chr	Pos(hg19)	Gene	Condon change	Aa change	Enzyme	Cut-site	5 reads
AD1	4	151835421	LRBA	Gca/Aca	A363T	Pst1	5'-CTGCA/G-3' 3'-G/ACGTC-5'	4/67 (6%)
AD2	2	189853334	COL3A1	Gga/Tga	G201*	ECO0191	5'-RG/GNCCY-3' 3'-YCCNG/GR-5'	4/42 (9%)

This table shows some of the additional somatic variations that we discovered in some hippocampus samples, but not in blood, using Virmid software. All of the mutations in this table are in genes that have hippocampus-specific SNVs, as indicated previously [6] and a restriction site for a restriction endonuclease, in such way that the presence of the mutation modifies that site. The table indicates the position of the mutation (according to hg19 assembly of the genome), the gene in which it is found, ID in dbSNP (none of them were previously identified in blood DNA samples), the amino-acid change caused, the enzyme that recognizes this site, and the percentage of reads with the mutation

rapid pathological autopsy shortly after death. The postmortem interval was 3 h. Immediately after the autopsy, the fresh tissues were flash-frozen in isopentane at -50°C . Thereafter, each frozen tissue sample was placed in a -80°C freezer for long-term preservation. Frozen tissue samples from various brain regions were obtained from the corresponding slices after 2-h of temperature soothing. Each sample was obtained with the aid of sterile disposable material and placed into sterile cryo-tubes. Thereafter, the samples were kept at -80°C . Blood samples were obtained simultaneously with routine blood extraction.

DNA isolation

All genomic DNA samples were isolated from blood or from the hippocampus, using Qiagen kits (DNeasy Blood and Tissue, ref: 69504) and following the manufacturer's instructions.

Sample processing for exome sequencing

3.10^{-6} g of genomic DNA was fragmented to an average size of 200 bp using a Covaris LE220 instrument. Short insert libraries were obtained using the Illumina TruSeq DNA Sample Preparation Kit. Exonic sequences were enriched using NimbleGen Sequence Capture Human Exome 2.1M Array. Paired-end sequences of 91 nucleotides from each end were generated to an average of 50x coverage using an Illumina HiSeq 2000 instrument. Sequences were generated in fastaq format.

Oligonucleotides used for DNA amplification (PCR)

The oligonucleotides listed in Table 2 were used for sequence amplification purposes. A concentration from 900 fM to 900 pM was used (see Fig. 6). DNA for PCR was purified with Wizard[®] Plus SV

Minipreps DNA Purification System (Ref A1330), following the manufacturer's guidelines.

Amplification of fragments bearing SNVs

For the experiment shown in Fig. 6, a modified version of TruePrime[™] technology from Sygnis[™] was used as an alternative to amplify and specifically enrich the samples containing the SNV of interest. TruePrime[™] is a multiple displacement amplification (MDA) technology based on the combination of Φ 29 DNA polymerase [16, 17] and TthPrimPol, a DNA primase that synthesizes DNA primers for Φ 29 DNA polymerase during the reaction, thus allowing the amplification of very low concentrations of DNA (Picher et al., submitted). The undigested DNA fragments (100 pg) containing the SNV were denatured (10 min at 95°C) and sealed (60 min at 60°C) to form circular single-stranded DNA molecules using Epicentre[®] CircLigase[™] II (100 units in a final reaction volume of 20 μl). Next, these circular DNA molecules were amplified by TruePrime[™] rolling circle amplification in the presence of the hexanucleotide (oligomer) (5'-ACCAAT-3') containing the motif for nuclease EC01091 (5'-ACCAAG-3'). In order to specifically enrich the amplification products with molecules bearing the SNV of interest, the G in this nuclease was changed to T—the SNV found in the AD patient. An oligonucleotide complementary to the reverse strand found in the AD patient was also added (5'-GGGTCA-3'). Briefly, the circularized DNA fragments (10 pg in 2.5 μl) were first denatured by adding 2.5 μl of buffer D and incubating 3 min at room temperature. The samples were then neutralized by adding 2.5 μl of buffer N. The amplification mix containing 26.8 μl of H₂O, 5 μl of reaction buffer, 5 μl of dNTPs, 5 μl of Enzyme 1 (TthPrimPol), and 0.7 μl of Enzyme 2 (ϕ 29DNApol) was added to the DNA samples, resulting in a final reaction volume of 50 μl . When indicated, 90 pM

Table 2
Oligonucleotides used for DNA amplification by PCR. f means forward and r reverse

Oligo	Seq. (5' → 3')	TM (°C)	Amplified region	Gene	SNV place
Or9g4.f	AACAAGGAGTCATGTGGTCAAT	56.5	Chr11 : 56510303-56511287	OR9G4	Chr11 : 56510623(A → G), rs513873, gTa/gCa (codon)
Or9g4.r COL3A1.f	CCGCCATGCAGCAATTTGTA TCAGCCATTCCAGCATGCATAA	56–63	Chr2 : 189852934-189853611	COL3A1	Chr2 : 189853334 (G → T) somatic mutation
COL3A1.r LRBA.f	AGCTCTGGGAAGTGGACTATT GTTACCATTTGCACACATCTC	52.5	Chr4 : 151,835,001-151,835,841	LRBA	Chr4 : 151835421(C → T), somatic mutation, Gca/Aca (codon)
LRBA.r	CTCAGTGGCATAGGATGCAG				

of each oligo (5'-ACCAAG-3' and 5'-GGGTCA-3') was added to the reaction. Reaction mixtures were incubated for 3 h at 30°C, and ϕ 29DNApol was inactivated for 10 min at 65°C to avoid degradation of the amplification products. The amount of amplified DNA obtained was approximately 1 μ g, as determined by the Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen, Life Technologies, Carlsbad, CA, USA). In some cases, the mixture was incubated (1 h) in the presence of Φ 29 DNA polymerase and the absence of TimePrime™ polymerase, followed by a 2-h incubation in the presence of both polymerases.

Sanger sequencing

Sanger sequencing for purified PCR products was performed using Applied Biosciences (ABI) 3720xl sequencers at GATC Biotech (Cologne, Germany).

Bioinformatics analysis

We proceeded in two different ways depending on whether our goal was to obtain well-defined SNVs with a homozygous or heterozygous genotype, wherein each is present 100% (if homozygous) or 50% (if heterozygous) of the cells, or whether the SNV is present in somatic mutations, where in each variation it is present at less than 50%.

Detection of SNVs

Samples were aligned to the human reference genome version hg19 [18] using the BWA aligner software [19] with default parameters. For each case, all the samples were pre-processed using Picard software to remove duplicate reads (<http://picard.sourceforge.net/>). Local realignment was performed around indels to improve SNV calling in these conflictive areas (IndelRealigner from the Genome Analyzer Toolkit, GATK, version 2.1–8 [20]). Base quality scores were recalibrated using BaseRecalibrator from GATK. The UnifiedGenotyper algorithm from GATK was then used with default parameters (see [21, 22] for details) to call SNVs, and a first file including raw calls was obtained. We then separated the indels from the rest of the calls and considered only SNVs for the analysis. These variants were filtered with VariantFiltration (from GATK) using the following parameters: coverage: DP > 10, DP > 20, DP > 50 or DP > 100, depending on the case of study; QD < 2.0; FS > 60.0, MQ < 35.0; HaplotypeScore > 13.0; MQRankSum

< -12.5 and ReadPosRankSum < -8.0. We selected only calls that passed these filters. Variants were annotated using the dbSNP database version 138 [23], the UCSC human RefGene [24], and the snpEFF software (version 2.0.5) [25]. In order to manipulate the files containing variations and to determine how many of these variations were unique or common to different tissues, we used the VCFtools software [26].

Detection of somatic mutations

To detect somatic mutations, we used hippocampus and blood samples from the same individual in each case and analyzed them together. We proceeded with the recalibration of the bases with BaseRecalibrator from GATK. After obtaining the recalibrated files, we used the Virmid software [15]. This algorithm is used for SNP profiling in paired control-disease samples with default parameters (taking, in our case, hippocampus as diseased tissue and blood as control). The mutations were obtained in VCF format and were also annotated using the snpEFF software (version 2.0.5) [25].

The symbols for the human genes, indicated in this work, suggested by HGNC are:

- LRBA (LPS responsive beige-like anchor protein) HGNC_ID: HGNC: 1742
- COL3A1 (Collagen type III alpha-1 chain) HGNC_ID: HGNC: 2201
- OR9G4 (Olfactory receptor family 9 subfamily G member 9) HGNC_ID: HGNC:15322

RESULTS

Validation of the analysis on a previously described SNV present in a high proportion of the tissue analyzed

As a first step, we used our method on a sample containing a previously described specific SNV. Using Genome Analysis Tool Kit (GATK)[20] software and following the recommended workflows for variant analysis proposed by the developers of this software, we previously identified among our samples [6] a SNV present in heterozygosis and in homozygosis (dbSNP ID: rs513873 A→G) in position chr11 : 56510623 (according to GRCh37 assembly of human genome) in *OR9G4* (olfactory receptor 9G4 gene, Fig. 1A). Using Virmid software, we confirmed the presence of the SNV (Fig. 1A) and then validated this finding by Sanger sequencing (Fig. 1B).

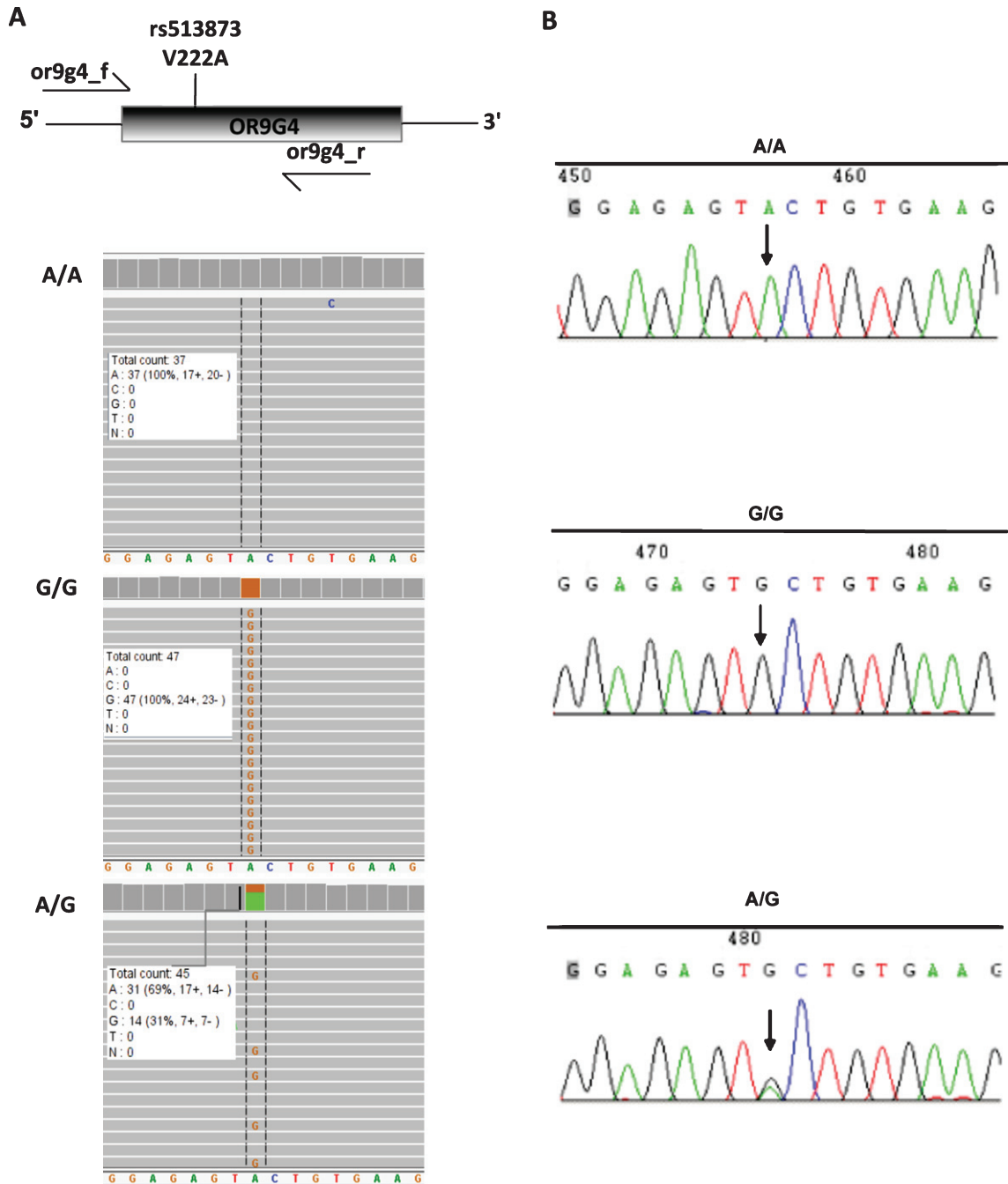


Fig. 1. Scheme showing the gene region where the SNV, determined by Illumina, is present. A) A single nucleotide variation in *or9g4r* gene, (dbSNP ID [2] : rs513873 A → G), in position chr11 : 56510623 (according to GRCh37 assembly of human genome) was detected. In order to test the efficiency of the Illumina-Virmid method, we sequenced three samples, with and without the variation, as described previously, in such way that A/A is homozygous for the reference allele and it does not have the SNV, G/G is homozygous for the described SNV, and A/G is heterozygous, having both the reference allele and the SNV respectively in each chromosome 11. The figure shows the alignments of the processed reads in the flanking regions of this SNV. As can be seen, each sequence shows a concrete number of reads with the reference allele and/or with the alternative allele (corresponding to the described SNV), thereby confirming its genotype (A/A, A/G, or G/G). B) This figure shows the chromatograms resulting from the Sanger sequencing analysis of the fragments containing (or not) the previously described SNV. The results obtained by Illumina sequencing described in Fig. 1A and the genotypes of the samples are corroborated here. The arrows indicate the exact site where the SNV is found in the sequence, pointing to a single peak for the homozygous samples (A/A and G/G) and a double peak for the heterozygous ones (A/G).

To amplify a fragment containing the SNV site, we designed two flanking DNA primers (see methods) to obtain a short DNA fragment (718 bp) that included this SNV. The single exon of *OR9G4* holds the AGTACT sequence, which, after the SNV, becomes AGTGCT. The first sequence (double stranded) 5'AGTACT3'; 3'TCATGA5' was cleaved by restriction enzyme ScaI, which cuts the 5'AGT/ACT3' site (Fig. 2A). After treatment of total DNA with the restriction enzyme, we found that only those DNA molecules lacking the SNV were digested. As a control, treatment with the enzyme of homozygous DNA for the SNV and for a control sample lacking it confirmed the previous results (Fig. 2B). This finding shows that a specific restriction nuclease can be used to cleave specific DNA sequences with or without the SNV.

Detection of low frequency brain-specific SNVs

We next tested whether the above procedure was useful to detect SNVs present less than 50% of the cells. As a proof of concept, we used the previously PCR-amplified DNA fragments of 718 bp corresponding to samples A/A (without the SNV and homozygous for the reference allele) and G/G (with the described SNV in both alleles and containing the target sequence for the restriction enzyme ScaI). We mixed these two DNA samples, which contain a different nucleotide in their sequences, at ratios of 1 : 1, 3 : 1, 5 : 1, and 10 : 1 (Fig. 3A) and repeated the procedure described above. The reference DNA sample (A), present in a higher number of cells, was a suitable target for the restriction nuclease ScaI (AGTACT), but not the DNA sample with the SNV (AGTGCT). To remove the reference DNA sequence, we digested the DNA mixture with ScaI. Next, we amplified and purified the undigested DNA and characterized the presence of the SNV by Sanger sequencing (Fig. 3B). Our results indicated that the method was effective to detect variants present in a 1/10 proportion. The main problem encountered was the low recovery yield of purified DNA after electrophoresis.

Identification of somatic brain mutations present in AD patients

We next sought to validate SNVs present in only a few cells of brain tissue. In a previous study on AD brains, we demonstrated the presence of somatic SNVs in this tissue. These were found to be absent in paired blood DNA samples. These analyses

were done by Illumina sequencing [6]. Here, using Virmid software, we validated two of these low frequency SNVs (Table 1). We also demonstrate that the presence of these SNVs prevented DNA digestion by specific nuclease restriction enzymes, thereby resulting in the removal of bulk DNA lacking the SNVs. These features may allow the use of the method shown in Supplementary Figure 1 for low frequency SNVs, detected by Illumina, by using standard Sanger sequencing.

From the SNVs shown in Table 1, we chose the exome of a patient with a SNV in the *COL3A1* gene in position chr2 : 189853334 G→T (see Fig. 4A). After the bioinformatics processing (Virmid), we found that this SNV was present in brain (9% of total reads) and absent in blood tissue. Direct Sanger sequencing of brain (Fig. 4B) and blood (Fig. 4C) samples showed the absence of this SNV in both blood and brain DNA. Since the presence of such a low frequency SNV prevents DNA digestion by nuclease Eco0109I, we digested brain DNA with this nuclease to remove bulk DNA.

Figure 5 shows the scheme, indicating that when the whole DNA (containing many molecules lacking the SNV and a few molecules containing it) was digested with Eco0109I, only those molecules of undigested DNA were isolated. These were later purified by gel electrophoresis, amplified by PCR, and sequenced by Sanger's method. However, we found that DNA recovery after gel electrophoresis was very low, thus hindering proper amplification of the residual DNA and the obtention of unequivocal data.

Specific amplification of fragments bearing the SNV of interest

First, a rolling circle replication approach was designed to amplify only the circularized fragments containing the SNV of interest. This was achieved by combining Φ 29 DNA polymerase with a specific hexanucleotide containing the nuclease motif for Eco0109I (5'-ACCAAG-3'), in which G was changed to T—the somatic SNV to be detected. However, the amount of DNA produced was insufficient for DNA sequencing using Sanger's method. Thus, we used an alternative approach (TruePrime™), based on Φ 29 DNA polymerase and *TthPrimPol* (Picher et al., submitted), to specifically and exponentially amplify this very low amount of DNA. The circularized DNA fragments containing the SNV were amplified by rolling circle replication using TruePrime™ (see Methods) in the presence of specific hexanucleotides,

A

>chr11:56510163-56510880_OR9G4

AACAAGGAGTCAT**GTGGTCAAT**ATTTTAATGTTTTAAATATG
 ACTTATTGAACTAATTGAACTACAGAAATGCAAATGAACAC
 ATTTATAAGCCAATAATCTGGAAAATCAATAACAGCATTG
 CAAAGAGAATAACCTTCATGTTTGTGGTTGTATAGTCTGTG
 TGCTTTCCTGAAGGCCTCTTTGATATCTTTGTTTCTCAGGCTA
 TAGATGAGAGGGTTGAGCAGTGGGTTGATCACGGTGTAGA
 ACAGAGCAGCTACTTTGTCCCTCTCTAGGGAGTAGGTGGAA
 CTAGGCCTTGAATACATAAAACAACAATGATCCATAGAAGAG
 CATGACTGAGATGAGGTGGGAAGCACAGGTGGAGAATGCC
 TTGTGCTTCTGAAGCTGAGTGGATTCTCAGGATAGCCAGG
 AGGATGTTGACATAGGAAATCAGGATAGCAAGAATGCTGG
 AGAGT**A**CTGTGAAGCCCACCACACCAAGCAGGACTTTTTCG
 TAGACCCTGGTGTGGTACAGGACATTTTACCAATGGTGGT
 GCATCACAGAAAAAGTGGTCAATGATATTTTACCACAAAA
 TGCAGGCGGAATGTATTGGCAGTATGGGCTATGGCATTCAA
 AAATCCTCCTATGTAGGAGCCAGCAACAAGCCCAGTACAGA
 GGGCGGTGGACATGGTACCTGAATAAAGCAATGGGT**TACA**
AATTGCTGCATGGCGG

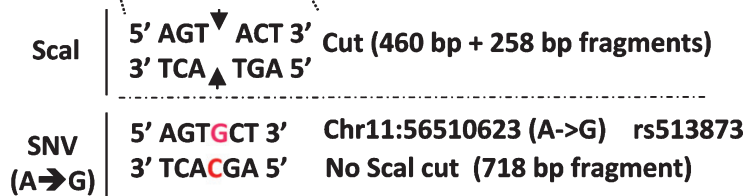
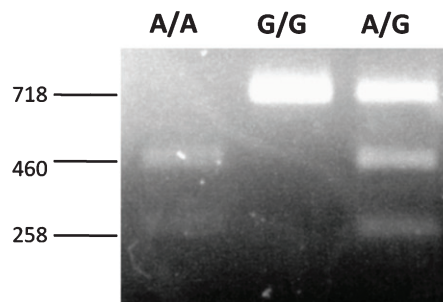
**B**

Fig. 2. Scheme showing the procedure developed to validate SNVs using restriction enzymes. A) The single nucleotide variation A→G was detected in such way that its presence modifies the recognition site for the restriction enzyme ScaI, thereby preventing its DNA cut. We designed two specific DNA primers to amplify a short region (718 bp) covering the site of this SNV in the genome. This fragment results in two small fragments (460bp + 258bp) when it does not contain the SNV or in a simple non-cleaved fragment (718 bp) when the SNV is present. B) This image shows the result of cutting the PCR-amplified fragments (see Fig. 1B) for the restriction enzyme ScaI. Samples were loaded in a 1% agarose gel after treatment with ScaI. Observe that, as indicated in 1B, the presence of the SNV prevents the enzyme from cutting the site. In A/A, the entire original 718 bp fragment has been digested, resulting in two fragments of 460 + 256 bp, G/G is homozygous for the SNV, so the cut has not occurred. Finally, in A/G, which is heterozygous for the SNV, there is a mix of the two previous results, with one half of its DNA digested and the other not.

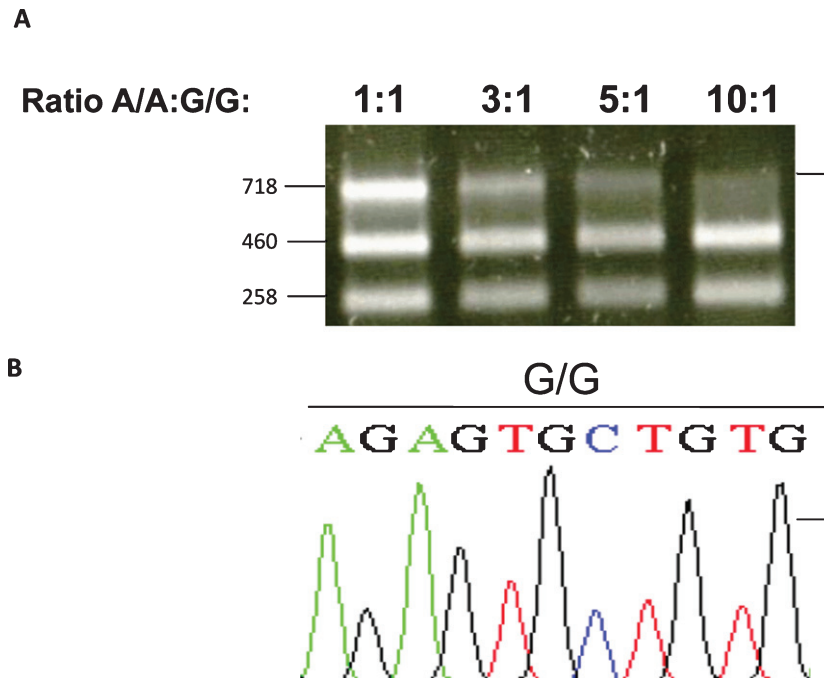


Fig. 3. Method to amplify a sample bearing a SNV that is present in a low proportion. A) Two samples containing a short 718 bp PCR-amplified DNA fragment having G/G or A/A, an SNV whose presence is crucial for the recognition of the restriction enzyme *ScaI* (see Fig. 1), were mixed in different ratios (1 : 1, 3 : 1, 5 : 1, and 10 : 1), in such way that the sample in a lower proportion was G/G, thus containing the SNV, and subsequently it was not recognized by the restriction enzyme *ScaI*. All the mixed samples were treated with this enzyme and loaded in an 1% agarose gel. It can be observed that the band at 718 bp matches proportionally in intensity the ratio of the sample of DNA with the SNV. B) The undigested band at 718 bp was recovered from the gel and sequenced by Sanger's method. Observe that this uncut band contains the described SNV.

in order to enrich the amplification products with molecules bearing the SNV of interest (Fig. 6A). We recovered at least 1 μg of amplified DNA per sample.

When this amplified DNA was sequenced by Sanger's method (Fig. 6C), we detected the described SNV, although it generally appeared together with some DNA lacking the variant. This observation could be attributed to the presence of a low proportion of undigested DNA lacking the SNV or to the presence of a low amount of DNA from other contaminant cells that result in a background of other nucleotide sequences. This background was more evident when brain tissue was analyzed, as we consistently achieved cleaner sequences from blood cells. Our results show that, by using specific DNA digestion, the above procedure allows the detection of low frequency SNVs (see also Figs. 1–3). In more detail, two variables for the amplification, namely oligonucleotide concentration and order for using (incubation) the $\Phi 29$ DNA polymerase and TruePrime™ polymerase, were taken into account. In the first case, increasing concentrations, from 900 fM to 900 pM of oligos, were used. In the second case,

all the components of the amplification mixture were added at the same time (method A) whereas for Method B, all the components of the amplification mixture (including $\Phi 29$ DNA polymerase), except TruePrime™ polymerase, were added for a first incubation (1 h – 30°C). Afterwards, TruePrime™ polymerase was added for a further incubation of the whole mixture (2 h – 30°C) (see Fig. 6A).

Figure 6B shows the results obtained for different samples amplified with distinct oligo concentrations and by using Method A or Method B. The (Sanger's) sequences obtained are shown in Fig. 6C.

In brain samples, SNVs present in less than 9–10% of total reads may be (Fig. 6) at the detection limit for validation by this method. Indeed, using the Virmid software, we observed the presence of specific SNVs in the other gene shown in Table 1 (see also Supplementary Figure 1). We tested the method described in Fig. 7 with the SNV found in position chr4:151835421 G→A of *LRB*. This SNV was present at a low frequency in brain (6% of total reads) and was not detected by standard Sanger sequencing (Supplementary Figure 2). This SNV was absent

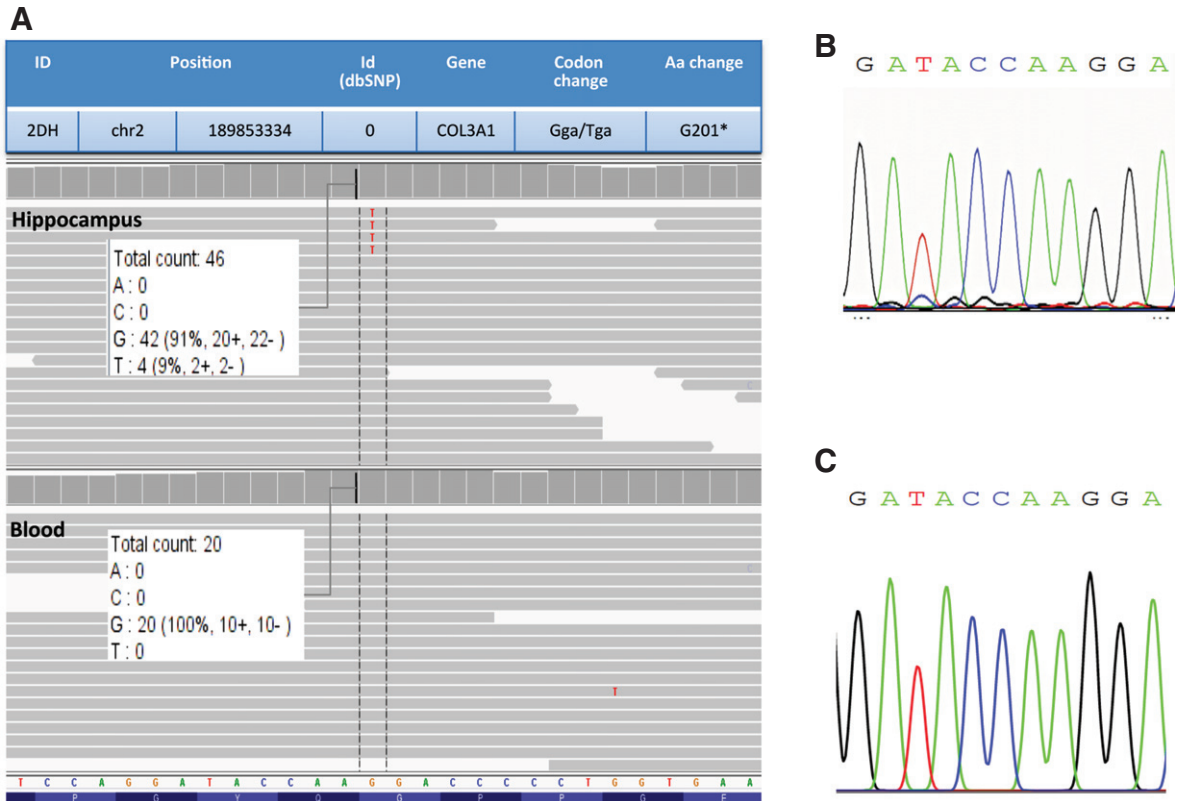


Fig. 4. Figure showing the alignments of the reads for the exomes of two different tissues (hippocampus and blood) from the same individual (2H). A) The alignments were detected for a somatic mutation by Virmid software (see methods). This site is in chromosome 2 in position 189853334 in *COL3A1* and results in the change G→T in that position. As can be seen, this variation is found in only 9% of the reads of the hippocampus, while it is absent in blood. Thus, brain and blood samples were sequenced by Sanger's method. B) Brain (hippocampal) DNA sample containing the SNV present in *COL3A1* in a low proportion was sequenced, but the SNV was not detected. C) Blood DNA samples from the same person as in (A), lacking the SNV present in *COL3A1*, were also sequenced. Curiously, cleaner sequences were obtained from blood than from brain samples.

in blood. The SNV was absent in the sequence 5'-CTGCAG-3', which was present in most of the DNA molecules and is the cleavage motif for restriction nuclease Pst1 (Table 1). Thus, we digested those molecules lacking the SNV with Pst1 and amplified the undigested DNA by PCR. However, since a very low amount of DNA was obtained in the PCR amplification, we were unable to obtain clear readouts after Sanger sequencing. In an attempt to achieve a better amplification, we used TruePrime™. We attained only a modest improvement of the data and found that the sequence with the specific SNV was contaminated by other sequences (Supplementary Figure 3). We thus conclude that the procedure described herein is suitable to validate low frequency allele variations present in at least 10% of DNA molecules by Sanger sequencing. Indeed, similar results to those of Supplementary Figure 3 were obtained when two

other genes, *wasf3* and *hsdl2*, were tested (data not shown).

DISCUSSION

Here we report a novel method (Fig. 7) to validate low frequency SNVs previously identified by Illumina sequencing in various tissues of a single donor. This method involves bioinformatics processing based on previously described methods [15]. Moreover, after removing the DNA lacking the somatic variations by means of digestion with specific restriction nucleases that can differentiate the presence of a single nucleotide, the remaining SNVs can be validated by Sanger's method. Although suitable nucleases are not always available for every SNV found, the good correlation between bioinformatics data and the cases validated by Sanger's method after

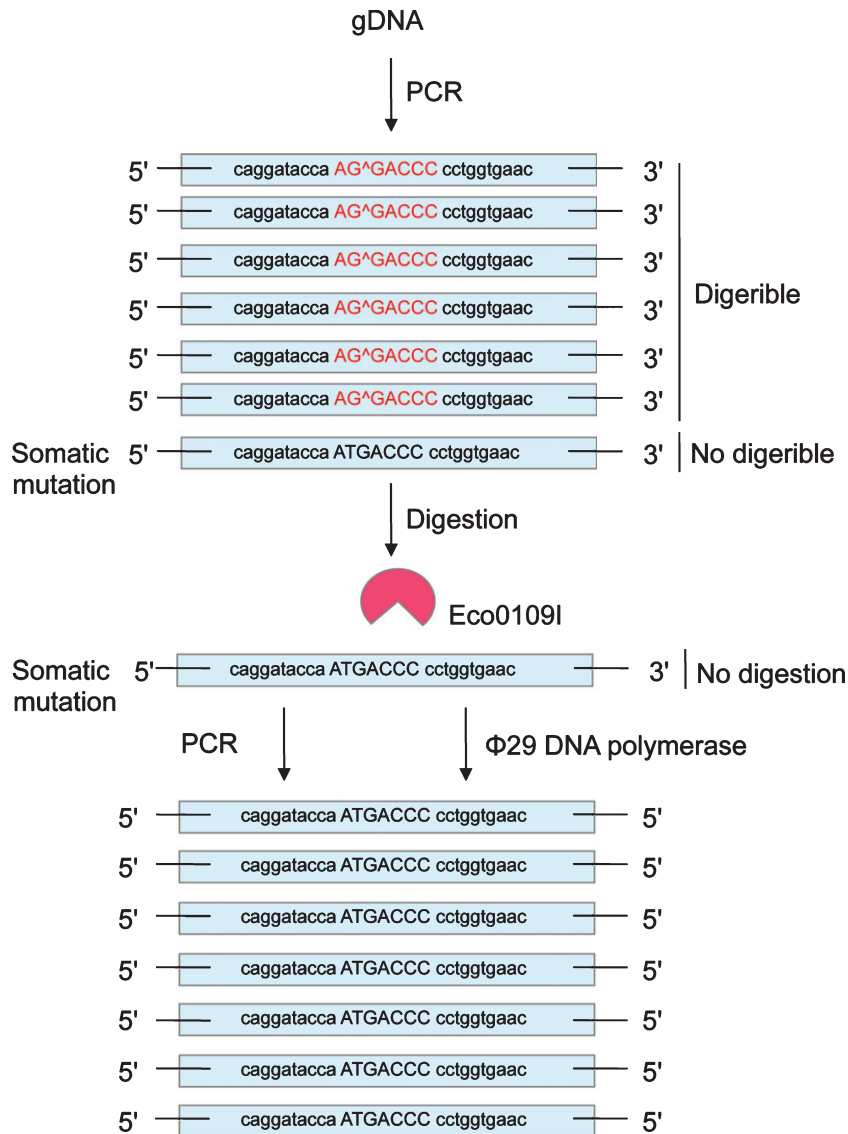


Fig. 5. Scheme showing the procedure developed to validate the presence of the SNV in hippocampus samples. As in the previous figure, the presence of the SNV modifies the cut site for the restriction enzyme Eco0109I, which recognizes the site AGGACCC present in the sequence. The variation modifies this sequence to ATGACCC, which is not recognized by the enzyme. Consequently, the sequence cannot be cut. This feature allowed us to recover the non-digested band from a gel and to amplify the sample by PCR to carry out posterior sequence analysis using Sanger's method.

nuclease treatment support the suitability and robustness of the bioinformatics procedures used to identify low-frequency variants.

As a first test for the use of restriction nucleases, we addressed whether we could fractionate DNA containing or lacking a given SNV in a donor with this SNV in heterozygosis. The gene containing the SNV, *OR9G4*, has been previously analyzed and the SNV reported (1000 genomes Project, <http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/>), the

allele frequencies for the Iberian population being $A = 0.7009$ and $G = 0.2991$. To look for specific a SNV present in less than 1/10 of the total reads, TruePrime™ technology (Picher et al., submitted) was used for rolling circle DNA amplification. In this technique, we used a specific oligomer primer whose motif contained the SNV.

Moreover, the presence of somatic mutations in single human neurons has recently been reported using single-cell sequencing [3, 4], being the main

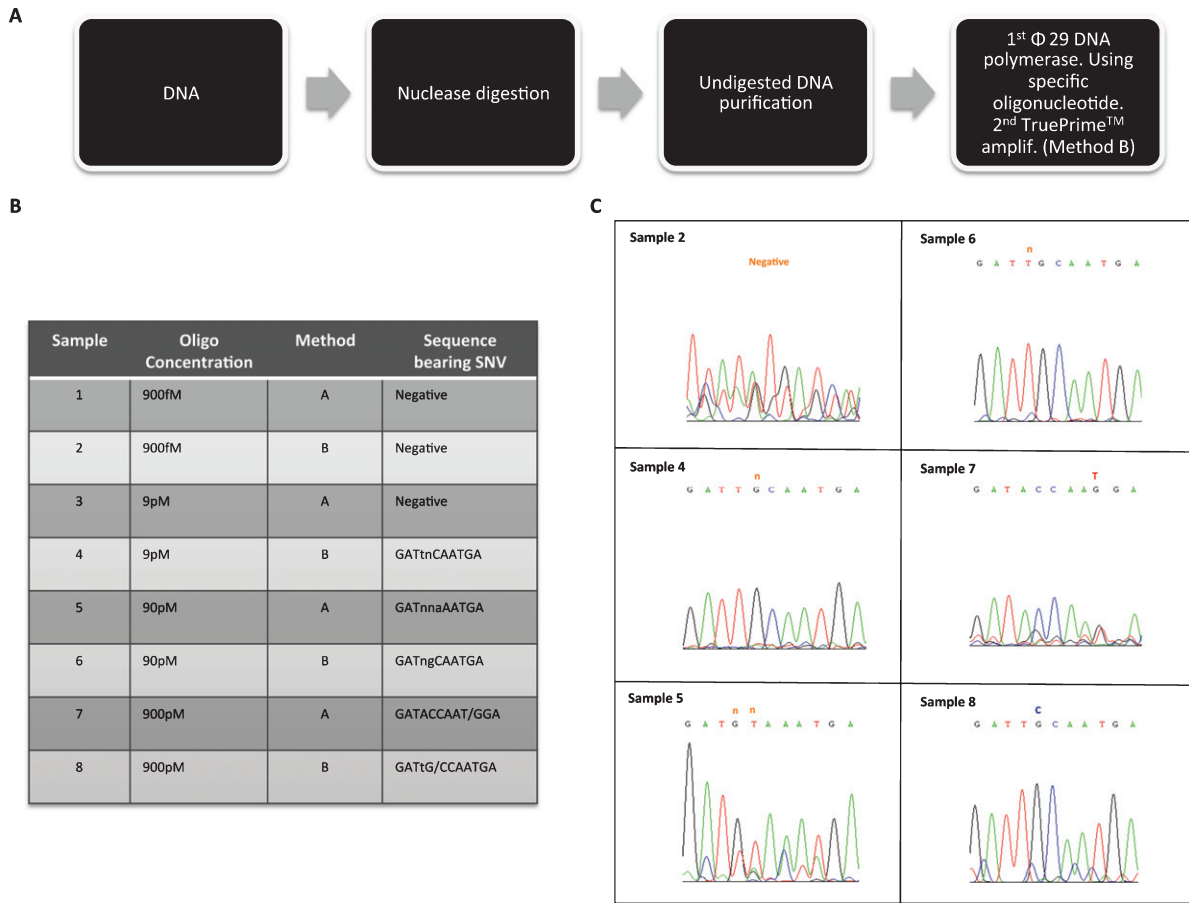


Fig. 6. Amplification of brain DNA containing the specific SNV by TruePrime™ and Sanger's sequencing of the same amplified sample. A) Scheme of circular DNA amplification by TruePrime™ using oligonucleotides containing the specific SNV. B) Effect of oligo concentration and method used for DNA amplification on obtaining the sequence bearing the SNV. C) Sequences, obtained by Sanger's, of samples 2 (negative), 4, 5, 6, 7, and 8.

cause for the appearance of SNVs, damage that could take place during transcription (see also the review [27]).

Also, it has been suggested that the presence of somatic SNVs induces the development of neurodevelopmental diseases [28]. In addition, new point mutations may occur at CpG dinucleotides, at cytosines that could be methylated [29]. Codons for arginine or glycine residues begin with a CG dinucleotide, which may explain why arginine or glycine is changed to other residues in a large proportion of non-synonymous somatic mutations [30]. Curiously, in one of our cases, we observed a change of a glycine residue (Table 1). Also, we found that none of the SNVs in Table 1 (AD patients) were described as single nucleotide polymorphisms in genome-wide association studies central data. However, the SNV (non-demented controls) shown in Fig. 1 was already

indicated in GWAS central data. The previous results are compatible with those SNVs shown in Table 1 arising by somatic mutations.

In relation to the genes analyzed in this study, *Or9g4* belongs to the olfactory receptor protein family and is a member of the G-protein-coupled receptors. The function of this member has not been analyzed. *Col3A1* encodes type III procollagen, and mutations cause type IV Ehlers-Danlos syndrome [31]. *LRBA* participates in intracellular vesicle trafficking and its expression is induced by lipopolysaccharide (LPS). Alterations in this gene are associated with a syndrome of immune deficiency and autoimmunity [32].

In summary, here we describe a method to identify low frequency somatic SNVs in non-proliferating cells like neurons. These variations can later be validated by Sanger sequencing. This method includes:

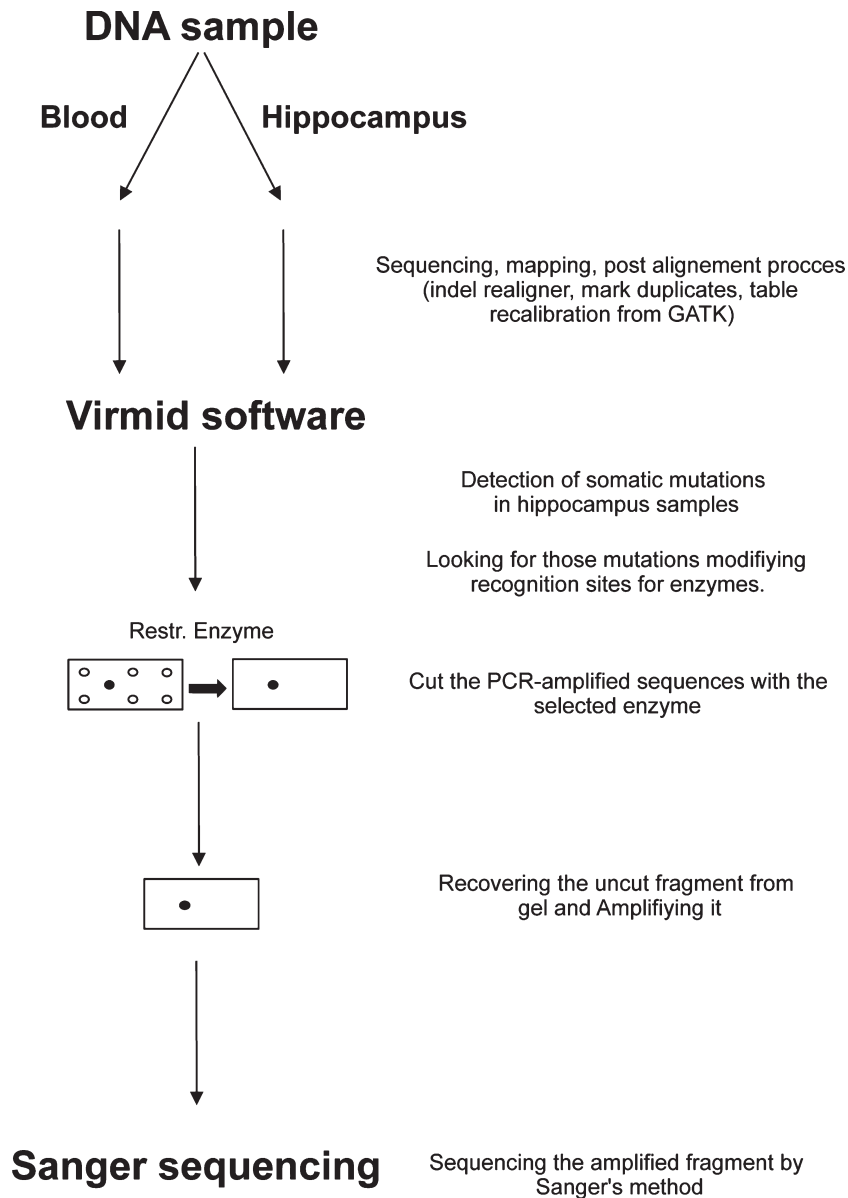


Fig. 7. Scheme showing the method developed in this work to detect somatic mutations that can be validated by Sanger's sequencing. After obtaining DNA samples from blood and hippocampus, their exomes were processed as indicated in materials and methods to obtain the reads in fastaq format in each case. A first process was made to obtain the sequences in bam format, using some of the tools included in GATK software [20]. These tools were: Indel realigned, Mark Duplicates and table recalibration (see methods) and the files were obtained in recal.bam form. After this, we processed the recal.bam files coming from blood and hippocampus samples with Virmid software, comparing hippocampal and blood samples in each case to obtain somatic mutations in hippocampus. The additional processes include looking for somatic mutations that are present in restriction nuclease motifs for a specific endonuclease. The use of that enzyme to enrich the sample having the SNV (due to the fact that its DNA can not be cut by the enzyme) and the recovery of this uncut sample from a gel and its amplification by PCR (when the SNV is present in at least 10% of the cells) or by using True Prime™ when a lower percentage of cells (or a very low amount of DNA) showing the SNV are present, may result in having a proper amount of DNA. Finally, sequencing by Sanger's method can be done.

a) the use of Virmid software; b) the removal of the DNA molecules lacking the somatic SNV by restriction nuclease digestion; and c) amplification of the DNA molecules containing the SNV of interest

by means of TruePrime™ technology and oligonucleotides containing the motif in which the SNV is present. The limitations of the method are that there is not always a restriction nuclease available to digest

DNA lacking the SNV of interest and that the minimal percentage of alternative reads that can be reliably validated by Sanger sequencing is 10% when brain tissue is used. Thus, the method proposed could be of help for validating low frequency tissue-specific mosaicism not only in the brain and in neurological diseases but in any tissue of interest. Furthermore, our data support the reliability of Virmid software to identify low frequency SNVs in Next-generation sequencing data.

ACKNOWLEDGMENTS

This work was supported mainly by grants from the *Fundación BBVA* (FBBVA) and the MICINN. We also acknowledge support from the *Fundación Reina Sofía* and grants from the *Plan Nacional* (MINECO) to JA (SAF-2014-53040-P) and ES (SAF-2013-4320). Human samples were kindly provided by Dr. Alberto Rabano (*Banco de Tejidos CIEN* [BT-CIEN]) and Dr. Rafael Sanchez (*Biobanco del Sistema Sanitario Público de Andalucía*). We would like to acknowledge the technical assistance of Ms. Nuria de la Torre.

Authors' disclosures available online (<http://j-alz.com/manuscript-disclosures/16-1053r2>).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: <http://dx.doi.org/10.3233/JAD-161053>.

REFERENCES

- Price DL, Tanzi RE, Borchelt DR, Sisodia SS (1998) Alzheimer's disease: Genetic studies and transgenic models. *Annu Rev Genet* **32**, 461-493.
- Tanzi RE (2012) The genetics of Alzheimer disease. *Cold Spring Harb Perspect Med* **2**.
- Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, Lee S, Chittenden TW, D'Gama AM, Cai X, Luquette LJ, Lee E, Park PJ, Walsh CA (2015) Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94-98.
- Hazen JL, Faust GG, Rodriguez AR, Ferguson WC, Shumilina S, Clark RA, Boland MJ, Martin G, Chubukov P, Tsunemoto RK, Torkamani A, Kupriyanov S, Hall IM, Baldwin KK (2016) The complete genome sequences, unique mutational spectra, and developmental potency of adult neurons revealed by cloning. *Neuron* **89**, 1223-1236.
- Sala Frigerio C, Lau P, Troakes C, Deramecourt V, Gele P, Van Loo P, Voet T, De Strooper B (2015) On the identification of low allele frequency mosaic mutations in the brains of Alzheimer's disease patients. *Alzheimers Dement* **11**, 1265-1276.
- Parcerisas A, Rubio SE, Muhaisen A, Gomez-Ramos A, Pujadas L, Puiggros M, Rossi D, Urena J, Burgaya F, Pascual M, Torrents D, Rabano A, Avila J, Soriano E (2014) Somatic signature of brain-specific single nucleotide variations in sporadic Alzheimer's disease. *J Alzheimers Dis* **42**, 1357-1382.
- Yates LR, Campbell PJ (2012) Evolution of the cancer genome. *Nat Rev Genet* **13**, 795-806.
- Tsiatis AC, Norris-Kirby A, Rich RG, Hafez MJ, Gocke CD, Eshleman JR, Murphy KM (2010) Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: Diagnostic and clinical implications. *J Mol Diagn* **12**, 425-432.
- Saksena G, Mermel C, Getz G (2012) Developing algorithms to discover novel cancer genes: A look at the challenges and approaches. *IEEE Signal Process Mag* **29**, 89-97.
- Gomez-Ramos A, Podlesniy P, Soriano E, Avila J (2015) Distinct X-chromosome SNVs from some sporadic AD samples. *Sci Rep* **5**, 18012.
- Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-219.
- Evrony GD, Lee E, Mehta BK, Benjamini Y, Johnson RM, Cai X, Yang L, Haseley P, Lehmann HS, Park PJ, Walsh CA (2015) Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49-59.
- Evrony GD, Lee E, Park PJ, Walsh CA (2016) Resolving rates of mutation in the brain using single-neuron genomics. *Elife* **5**, pii: e12966.
- Poduri A, Evrony GD, Cai X, Elhosary PC, Beroukhir R, Lehtinen MK, Hills LB, Heinzen EL, Hill A, Hill RS, Barry BJ, Bourgeois BF, Riviello JJ, Barkovich AJ, Black PM, Ligon KL, Walsh CA (2012) Somatic activation of AKT3 causes hemispheric developmental brain malformations. *Neuron* **74**, 41-48.
- Kim S, Jeong K, Bhutani K, Lee J, Patel A, Scott E, Nam H, Lee H, Gleeson JG, Bafna V (2013) Virmid: Accurate detection of somatic mutations with sample impurity inference. *Genome Biol* **14**, R90.
- Blanco L, Bernad A, Lazaro JM, Martin G, Garmendia C, Salas M (1989) Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J Biol Chem* **264**, 8935-8940.
- Dean FB, Nelson JR, Giesler TL, Lasken RS (2001) Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res* **11**, 1095-1099.
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford K, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT,

- Pepin KH, Gish WR, Chisoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brothier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowski J, International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921.
- [19] Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595.
- [20] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303.
- [21] DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-498.
- [22] Pagnamenta AT, Lise S, Harrison V, Stewart H, Jayawant S, Quaghebeur G, Deng AT, Murphy VE, Sadighi Akha E, Rimmer A, Mathieson I, Knight SJ, Kini U, Taylor JC, Keays DA (2012) Exome sequencing can detect pathogenic mosaic mutations present at low allele frequencies. *J Hum Genet* **57**, 70-72.
- [23] Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K (2001) dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res* **29**, 308-311.
- [24] Dreszer TR, Karolchik D, Zweig AS, Hinrichs AS, Raney BJ, Kuhn RM, Meyer LR, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B, Pohl A, Malladi VS, Li CH, Learned K, Kirkup V, Hsu F, Harte RA, Guruvadoo L, Goldman M, Gardine BM, Fujita PA, Diekhans M, Cline MS, Clawson H, Barber GP, Haussler D, James Kent W (2012) The UCSC Genome Browser database: Extensions and updates 2011. *Nucleic Acids Res* **40**, D918-923.
- [25] Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92.
- [26] Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis Group (2011) The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158.
- [27] Avila J, Gomez-Ramos A, Soriano E (2014) Variations in brain DNA. *Front Aging Neurosci* **6**, 323.
- [28] Veltman JA, Brunner HG (2012) De novo mutations in human genetic disease. *Nat Rev Genet* **13**, 565-575.
- [29] Shendure J, Akey JM (2015) The origins, determinants, and consequences of human mutations. *Science* **349**, 1478-1483.
- [30] Cooper DN, Krawczak M (1990) The mutational spectrum of single base-pair substitutions causing human genetic disease: Patterns and predictions. *Hum Genet* **85**, 55-74.
- [31] Liu X, Wu H, Byrne M, Krane S, Jaenisch R (1997) Type III collagen is crucial for collagen I fibrillogenesis and for normal cardiovascular development. *Proc Natl Acad Sci U S A* **94**, 1852-1856.
- [32] Lopez-Herrera G, Tampella G, Pan-Hammarstrom Q, Herholz P, Trujillo-Vargas CM, Phadwal K, Simon AK, Moutschen M, Etzioni A, Mory A, Srugo I, Melamed D, Hultenby K, Liu C, Baronio M, Vitali M, Philippet P, Dideberg V, Aghamohammadi A, Rezaei N, Enright V, Du L, Salzer U, Eibel H, Pfeifer D, Veelken H, Stauss H, Lougaris V, Plebani A, Gertz EM, Schaffer AA, Hammarstrom L, Grimbacher B (2012) Deleterious mutations in LRBA are associated with a syndrome of immune deficiency and autoimmunity. *Am J Hum Genet* **90**, 986-1001.