ANNALS of Clinical and Translational Neurology
Open Access
ANA American Neurological Association since 1875

RESEARCH ARTICLE

Archetypal *NOTCH3* mutations frequent in public exome: implications for CADASIL

Julie W. Rutten[1,2], Hans G. Dauwerse[1,2], Gido Gravesteijn[1], Martine J. van Belzen[1], Jeroen van der Grond[3], James M. Polke[4], Manuel Bernal-Quiros[4] & Saskia A. J. Lesnik Oberstein[1]

[1]Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands
[2]Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
[3]Department of Radiology, Leiden University Medical Center, Leiden, The Netherlands
[4]Neurogenetics Unit, National Hospital for Neurology and Neurosurgery, London, United Kingdom

Correspondence
Saskia A.J. Lesnik Oberstein, Clinical Geneticist. Department of Clinical Genetics, K5-R Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, The Netherlands. Tel: 00-31-(0)71-5268033; Fax: 00-31-(0)71-5266749; E-mail: lesnik@lumc.nl

Funding Information
This work was financially supported by ZonMw, The Netherlands Organization for Health Research and Development, grant no. 40-41900-98-018.

Received: 14 June 2016; Revised: 1 August 2016; Accepted: 1 August 2016

Annals of Clinical and Translational Neurology **2016; 3(11): 844–853**

doi: 10.1002/acn3.344

Abstract

Objective: To determine the frequency of distinctive EGFr cysteine altering *NOTCH3* mutations in the 60,706 exomes of the exome aggregation consortium (ExAC) database. **Methods**: ExAC was queried for mutations distinctive for cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL), namely mutations leading to a cysteine amino acid change in one of the 34 EGFr domains of NOTCH3. The genotype-phenotype correlation predicted by the ExAC data was tested in an independent cohort of Dutch CADASIL patients using quantified MRI lesions. The Dutch CADASIL registry was probed for paucisymptomatic individuals older than 70 years. **Results**: We identified 206 EGFr cysteine altering *NOTCH3* mutations in ExAC, with a total prevalence of 3.4/1000. More than half of the distinct mutations have been previously reported in CADASIL patients. Despite the clear overlap, the mutation distribution in ExAC differs from that in reported CADASIL patients, as mutations in ExAC are predominantly located outside of EGFr domains 1–6. In an independent Dutch CADASIL cohort, we found that patients with a mutation in EGFr domains 7–34 have a significantly lower MRI lesion load than patients with a mutation in EGFr domains 1–6. **Interpretation**: The frequency of EGFr cysteine altering *NOTCH3* mutations is 100-fold higher than expected based on estimates of CADASIL prevalence. This challenges the current CADASIL disease paradigm, and suggests that certain mutations may more frequently cause a much milder phenotype, which may even go clinically unrecognized. Our data suggest that individuals with a mutation located in EGFr domains 1–6 are predisposed to the more severe "classical" CADASIL phenotype, whereas individuals with a mutation outside of EGFr domains 1–6 can remain paucisymptomatic well into their eighth decade.

Introduction

Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) is a monogenic small vessel disease caused by highly distinctive mutations in the *NOTCH3* gene.[1] *NOTCH3* mutations in CADASIL lead to (cerebro)vascular NOTCH3 protein aggregation, compromised cerebral blood flow, mid-adult onset of stroke and vascular cognitive impairment, and migraine with aura. Although most patients experience their first stroke at 45–50 years of age, there is a wide variability in age of onset and disease progression,[2] and a later-onset, milder disease course is increasingly recognized.[3]

CADASIL patients have progressive ischemic brain MRI abnormalities, which correlate with disease severity.[4–6] In the presymptomatic stage of the disease, up to decades before onset of symptoms, symmetric periventricular white matter hyperintensities (WMH) are observed,[7,8] often also affecting the anterior temporal

lobes. The MRI in symptomatic individuals often also reveals multiple lacunar infarcts, microbleeds, and brain atrophy.[9,10]

CADASIL has been reported worldwide, but only some countries in Europe and East Asia have compiled large series of CADASIL patients and their *NOTCH3* mutations.[11–17] Prevalence studies have been performed in relatively small populations, with a reported minimum prevalence of 2-5/100,000.[3,18,19] This number is widely held to be an underestimation.[3]

Over 98% of CADASIL patients have a distinctive missense mutation in one of *NOTCH3* exons 2–24, invariably leading to the gain or loss of a cysteine residue in one of the 34 epidermal growth factor-like repeat (EGFr) domains of the NOTCH3 protein. This changes the highly conserved number of six cysteines in one of these EGFr domains to an uneven number of five or seven cysteines.[12] The resultant unpaired cysteine is predicted to lead to abnormal disulphide bridge formation and NOTCH3 protein aggregation.[20] More than 200 distinct *NOTCH3* mutations have been identified in CADASIL patients, dispersed across EGFr domains 1–34, but with the highest prevalence in EGFr domains 1–6.[21] Nonpenetrance has never been reported for this distinct type of *NOTCH3* mutation. A definite, clinically relevant genotype-phenotype correlation has never been found.[13,14]

In this study, we determined the prevalence of EGFr cysteine altering *NOTCH3* mutations in the large publically accessible exome sequencing dataset ExAC and performed comparative analyses with the over 200 *NOTCH3* mutations reported in both the Dutch and international CADASIL population. The genotype-phenotype association predicted by comparing ExAC and CADASIL mutation spectra was assessed in an independent cohort of Dutch CADASIL patients. We discuss the implications of our findings for the prevalence of CADASIL, the ever-broadening disease spectrum, and the discovery of a novel genotype-phenotype correlation.

Materials and Methods

The Exome Aggregation Consortium dataset

NOTCH3 mutation frequencies were assessed using the browser of the Exome Aggregation Consortium (ExAC) (Lek et al.; http://dx.doi.org/10.1101/030338; http://exac.broadinstitute.org; accessed Oct/Nov 2015). This publically accessible database contains exome data of 60,706 unrelated individuals; 33,644 males and 27,062 females. The exome data are compiled from 14 independent disease and control cohorts. No specific stroke or dementia cohorts are included in ExAC. Individuals in ExAC are assigned to one of six populations: European (n = 36,677; Finnish n = 3307, Non-Finnish n = 33,370), Latino (n = 5789), African/African American (n = 5203), South Asian (n = 8256), East Asian (n = 4327), and Other (n = 454). Data in ExAC are based on genome build GRCh37/hg19.

NOTCH3 variant ascertainment in ExAC

We queried ExAC for missense mutations in *NOTCH3* that passed all ExAC quality checks (Lek et al.; http://dx.doi.org/10.1101/030338). Next, these variants were filtered using stringent criteria for pathogenicity in CADASIL, namely missense mutations leading to the gain or the loss of a cysteine residue in one of the 34 EGFr domains of the *NOTCH3* protein (amino acid residues 40–1373) (http://www.uniprot.org). Mean coverage of the *NOTCH3* canonical transcript in ExAC is 35.72×. Most of the EGFr encoding exons (exons 2–24) have a coverage of at least 10×, except for exon 15 and a part of exon 24, which have a coverage below 10x. Four independently accessible cohorts contributing to the ExAC database were also individually queried for pathogenic *NOTCH3* mutations: 1000 Genomes (http://www.1000genomes.org); NHLBI-GO Exome Sequencing Project (http://evs.gs.washington.edu); Sequencing Initiative Suomi (http://www.sisuproject.fi/) and Type 2 DiabetesGenetics 17k exome sequence analysis (http://www.type2diabetesgenetics.org).

Comparison of *NOTCH3* mutations in ExAC to those reported in CADASIL patients

NOTCH3 mutations present in ExAC were compared to those reported in the Dutch CADASIL registry and the international CADASIL literature. To avoid bias through incomplete reporting of mutation frequencies or founder effects, the comparative analysis of the mutation distribution across the 34 EGFr domains was performed using distinct mutations rather than the prevalence of these mutations (i.e. each mutation was included only once in all analyses, independent of its total frequency in the CADASIL population). Next to a comparison with the mutation spectrum reported worldwide, we also separately compared the spectrum of ExAC mutations to those in the Dutch registry alone because, in contrast to many other countries, the *NOTCH3* mutation spectrum in the Dutch population is well characterized, as *NOTCH3* mutation screening in the Netherlands is comprehensive for exons 2–24, has been operational for more than 15 years and molecular diagnostics are readily accessible to all Dutch citizens. Moreover, all Dutch *NOTCH3* mutations are registered in a single national registry, which includes 383 individuals with 45 distinct mutations from 163 families. For mutation annotation, the reference

sequence NM_000435.2 was used and sequence variants were described according to the HGVS nomenclature recommendations.

Genotype-phenotype analysis

The confirmation of an EGFr- location-dependent genotype-phenotype correlation was performed in an independent CADASIL cohort. A detailed characterization of this cohort has been described elsewhere.[10] A standardized MRI protocol was performed at a field strength of 1.5T (Philips Medical System). Image acquisition and volumetric analysis of WMH, brain parenchyma, and lacunar infarcts were performed as previously described.[4] Briefly, WMHs were defined as white matter areas with increased signal intensity on T2-w MRI and FLAIR. WMH volume was calculated as a percentage of the brain parenchyma volume. Lacunar infarcts were defined as parenchymal defects not extending to the cortical grey matter, with a signal intensity corresponding to that of cerebrospinal fluid and having a diameter between 2 mm and 1.5 cm. Statistical analysis was performed using the SPSS statistical software package (version 23) (SPSS Inc., Chicago). Comparison of quantitative brain MRI variables was performed using linear regression, co- varying for age, sex, smoking, and hypertension.

NOTCH3 phenotypic spectrum in over 70 year olds

As ExAC genotypes cannot be traced back to the phenotypic data of individuals, we could not assess the correlation of genotype with (possible) CADASIL phenotype in ExAC. To determine whether *NOTCH3* mutations can be present in asymptomatic or paucisympomatic individuals and whether nonpenetrance may occur, we queried our Dutch CADASIL registry for individuals of at least 70 years of age at the time of DNA-testing. We selected only those individuals with an available brain MRI or CT scan, determined the reason for *NOTCH3* mutation testing, and reviewed their clinical records and brain scans.

Results

High frequency of EGFr cysteine altering *NOTCH3* mutations in ExAC

We found a total number of 642 *NOTCH3* missense variants in ExAC, of which 206 fit all the criteria for pathogenicity in CADASIL, namely mutations leading to an uneven number of 5 or 7 cysteine amino acids in one of the 34 EGFr domains of the *NOTCH3* protein (Table 1). The frequency of these mutations in ExAC is therefore 206 in 60,706 exomes, that is 3.4/1000. This is 100-fold higher than what would be expected based on the current CADASIL prevalence estimations of 2–5/100,000.[3,18,19] In the four independently accessible ExAC cohorts, we found the following frequencies of pathogenic *NOTCH3* mutations: Type 2 diabetes cohort (n = 16,857): 3.9/1000; NHLBI-GO Exome Sequencing Project (n = 6503): 1.2/1000; Sequencing Initiative Suomi (n = 6118): 0.7/1000; 1000 Genomes (n = 2504): 3.6/1000. This confirmation of high mutation frequencies in all four independently accessible cohorts shows that the high *NOTCH3* mutation frequency in ExAC is not attributable to a specific ExAC cohort.

Characteristics of *NOTCH3* mutations in ExAC

The 206 EGFr cysteine altering missense mutations in ExAC consist of varying frequencies of 25 distinct mutations (Table 1). Fourteen of the 25 distinct mutations are present in ExAC only once. The other mutations are recurrent, and are present between two and eight times, except for two mutations which have high frequencies of 32× and 120×. These two mutations are found across ethnicities, but are most frequent in the East Asian (p.Arg544Cys) and South Asian (p.Arg1231Cys) populations (Fig. 1). The most frequent mutation (p.Arg1231Cys) is present in a homozygous state in three individuals, in a population reported to be highly consanguineous (Lek et al.; http://dx.doi.org/10.1101/030338). The mutations are located in 10 different *NOTCH3* EGFr encoding exons, encoding a total of 12 EGFr domains. Exons 11 and 22 harbor the most distinct mutations, namely six each, followed by exons 4 and 12 with three distinct mutations each. Overall, 17 of the 25 mutations lead to the gain of a cysteine residue in an EGFr domain and eight to the loss of a cysteine.

Comparison of *NOTCH3* mutations in ExAC to those reported in CADASIL patients

We found pathogenic *NOTCH3* mutations in each of the populations included in ExAC (Fig. 1). Mutation frequencies, however, differ between the populations and range from 0.4/1000 in African/African Americans to 11.7/1000 in South Asians. Some mutations are present in many different populations; the p.Arg578Cys mutation and the p.Arg1231Cys mutation, for example, are each found in four of six populations. Mutations found in specific ethnic populations in ExAC have also been reported in CADASIL patients from the same ethnicity or geographical region. For example, the p.Arg544Cys mutation which is very frequent in East Asians in ExAC, has also been described frequently in CADASIL patients from Taiwan[22]

Table 1. Twenty-five distinct cysteine altering *NOTCH3* mutations in the ExAC database.

cDNA	Protein	Exon	EGFr	Count in ExAC	Mutation previously reported in CADASIL
c.350G>T	**p.Cys117Phe**	**4**	**2**	**1**	**yes[32]**
c.619C>T	**p.Arg207Cys**	**4**	**5**	**1**	**yes[33]**
c.635G>A	**p.Cys212Tyr**	**4**	**5**	**1**	**yes[34]**
c.931T>G	p.Cys311Gly	6	7	1	No
c.1045T>A	p.Cys349Ser	7	8	1	No
c.1630C>T	**p.Arg544Cys**	**11**	**14**	**32**	**yes[33]**
c.1672C>T	**p.Arg558Cys**	**11**	**14**	**5**	**yes[12]**
c.1732C>T	**p.Arg578Cys**	**11**	**14**	**5**	**yes[12]**
c.1759C>T	**p.Arg587Cys**	**11**	**15**	**2**	**yes[35]**
c.1819C>T	**p.Arg607Cys**	**11**	**15**	**1**	**yes[36]**
c.1823G>A	p.Cys608Tyr	11	15	1	No
c.1871G>C	p.Cys624Ser	12	16	1	No
c.1903C>T	p.Arg635Cys	12	16	1	No
c.1918C>T	**p.Arg640Cys**	**12**	**16**	**3**	**yes [LOVD]**
c.2149C>T	**p.Arg717Cys**	**14**	**18**	**3**	**yes [HGMD]**
c.2824G>T	p.Gly942Cys	18	24	1	No
c.2851C>T	p.Arg951Cys	18	24	1	No
c.3427C>T	**p.Arg1143Cys**	**21**	**29**	**2**	**yes [HGMD]**
c.3568C>T	p.Arg1190Cys	22	30	5	No
c.3601C>T	p.Arg1201Cys	22	30	7	No
c.3605G>C	p.Cys1202Ser	22	30	1	No
c.3628C>T	**p.Arg1210Cys**	**22**	**31**	**1**	**yes [Netherlands]**
c.3664T>G	**p.Cys1222Gly**	**22**	**31**	**8**	**yes[3]**
c.3691C>T[1]	**p.Arg1231Cys[1]**	**22**	**31**	**120**	**yes[12]**
c.3724C>T	p.Arg1242Cys	23	31	1	No
Total				**206**	

Fourteen of the 25 distinct missense mutations have been previously described in CADASIL patients. The others all fit the criteria for pathogenicity in CADASIL, namely leading to a cysteine amino acid change in one of the 34 EGFr domains of the NOTCH3 protein. Previously reported *NOTCH3* mutations in CADASIL patients are shown in bold.

CADASIL, cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy; LOVD, Leiden Open Variation Database (www.lovd.nl/NOTCH3); HGMD, The Human Gene Mutation Database (http://www.hgmd.cf.ac.uk).

[1]Mutation reported three times in a homozygous state.

and from Korea.[11] Likewise, the p.Arg1231Cys mutation, which is the most frequent mutation in Europeans in ExAC, has also been reported in three of the large European CADASIL cohorts.[12,13,15]

Mutations in ExAC are predominantly located outside of EGFr domains 1-6

Next, we compared the distribution of mutations in ExAC to those in the well-characterized Dutch CADASIL registry and to those reported in CADASIL patients worldwide. In the Dutch CADASIL registry, there are 45 distinct cysteine altering *NOTCH3* mutations, of which six have not been previously reported (Table S1). The distribution of Dutch CADASIL-causing *NOTCH3* mutations largely overlaps with those reported worldwide,[21] with the largest percentage of distinct mutations located in exon 4 (Fig. 2). In ExAC, the percentage of distinct mutations in exon 4 is three times lower than in the Dutch CADASIL registry (12% vs. 38%). Conversely, the percentage of distinct mutations in exon 22 is 10 times higher in ExAC than in the Dutch CADASIL registry (24% vs. 2%). There is a relatively high percentage of distinct mutations in exon 11 in both ExAC and the Dutch CADASIL registry (24% and 16%, respectively). Of note, although the distribution of mutations in ExAC differs from the distribution found in reported CADASIL patients, the majority of the exact same distinct EGFr cysteine altering *NOTCH3* mutations in ExAC, namely 14 out of 25 (56%), have been previously reported in CADASIL patients (Table 1). Finally, we compared the mutation spectrum at the protein level using the distribution across the 34 *NOTCH3* EGFr domains. This shows that mutations in ExAC are mostly clustered in EGFr domains 14–16 and 29–31, whereas reported CADASIL mutations cluster in EGFr domains 1–6.

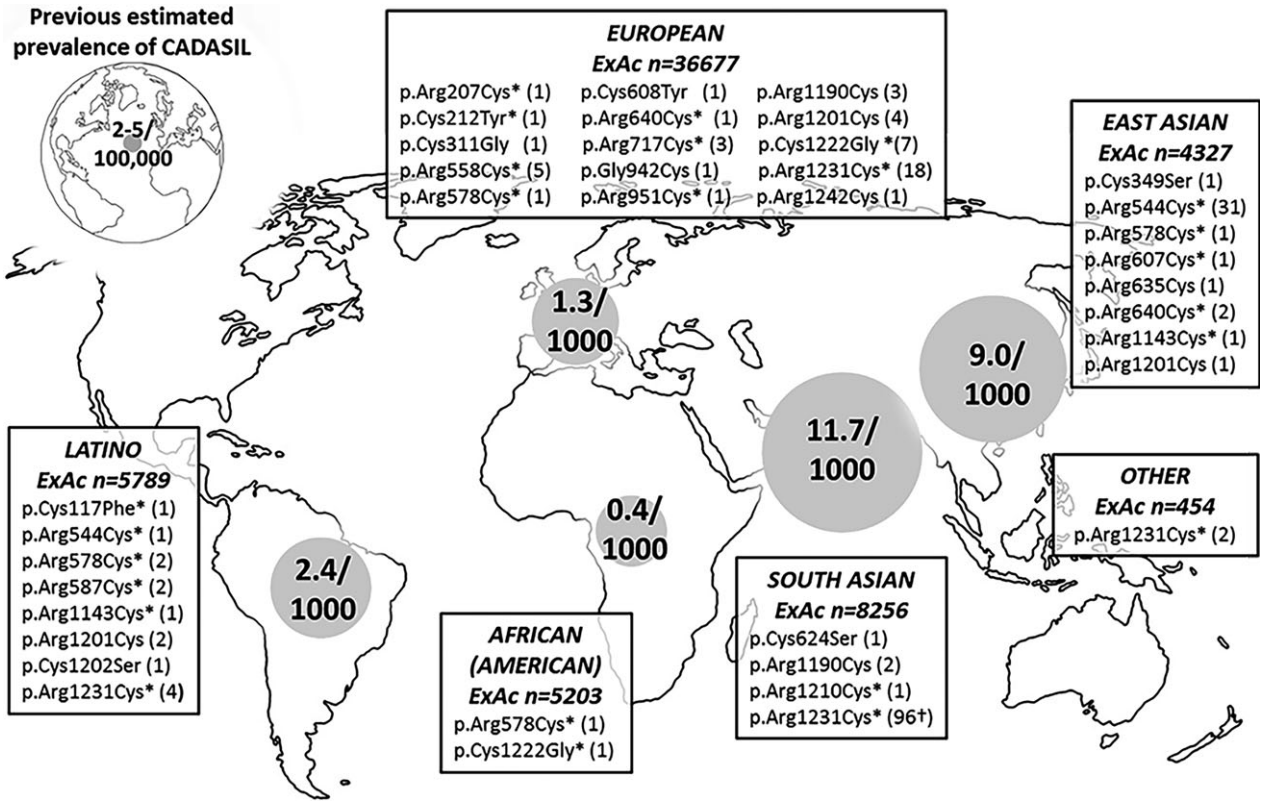

Figure 1. High frequency of EGFr cysteine altering *NOTCH3* mutations in all populations in the ExAC database. Frequencies range from 0.4/1000 individuals in the African/African American population to 11.7/1000 in the South Asian population, with an overall frequency of 3.4/1000. *: Mutation previously described in cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) patients; †: Mutation reported three times in a homozygous state; (): Allele counts

NOTCH3 mutations outside of EGFr domains 1–6 are correlated with a lower MRI lesion load

The observed difference in mutation distribution between mutations in ExAC and those reported in CADASIL led us to hypothesize that a genotype-phenotype correlation may play a role. Specifically, we hypothesized that mutations outside of EGFr domains 1–6, the classical CADASIL predominance region, predispose to a less severe phenotype. To test this hypothesis, we compared the WMH lesion load and number of lacunar infarcts between patients with a mutation in EGFr domains 1–6 to patients with a mutation in EGFr domains 7–34. We found that patients with mutations in EGFr domains 1–6 had a significantly higher WMH lesion load than patients with mutations in EGFr domains 7–34 (Fig. 3A). A similar trend was seen for the number of lacunar infarcts, almost reaching statistical significance (Fig. 3B). Although group size for this analysis is relatively small, these results suggest that mutations outside of *NOTCH3* EGFr domains 1–6 are associated with a less severe brain MRI phenotype.

NOTCH3 mutations are penetrant on brain imaging in paucisymptomatic elderly individuals

To determine whether *NOTCH3* mutations can be associated with nonpenetrance or an undiagnosed phenotype, possibly explaining the high ExAC mutation frequency, we selected individuals from the Dutch CADASIL registry who were 70 years or older at the time of DNA-testing and in whom a recent brain scan was available. We identified four such individuals. The reasons for DNA-testing were either predictive testing because of the diagnosis of CADASIL in a family member, or because of a chance finding of WMHs on brain imaging, performed for unrelated signs or symptoms, such as hearing loss. The mutations in these patients were in exon 10 (p.Cys531Gly) and in exon 11 (p.Arg587Cys). None of them had a history of stroke. One patient reported mild complaints of memory loss, the other three had no cognitive complaints at all and no cognitive deficits were reported by their partners. All individuals lived at home and were fully independent for activities of daily living. One

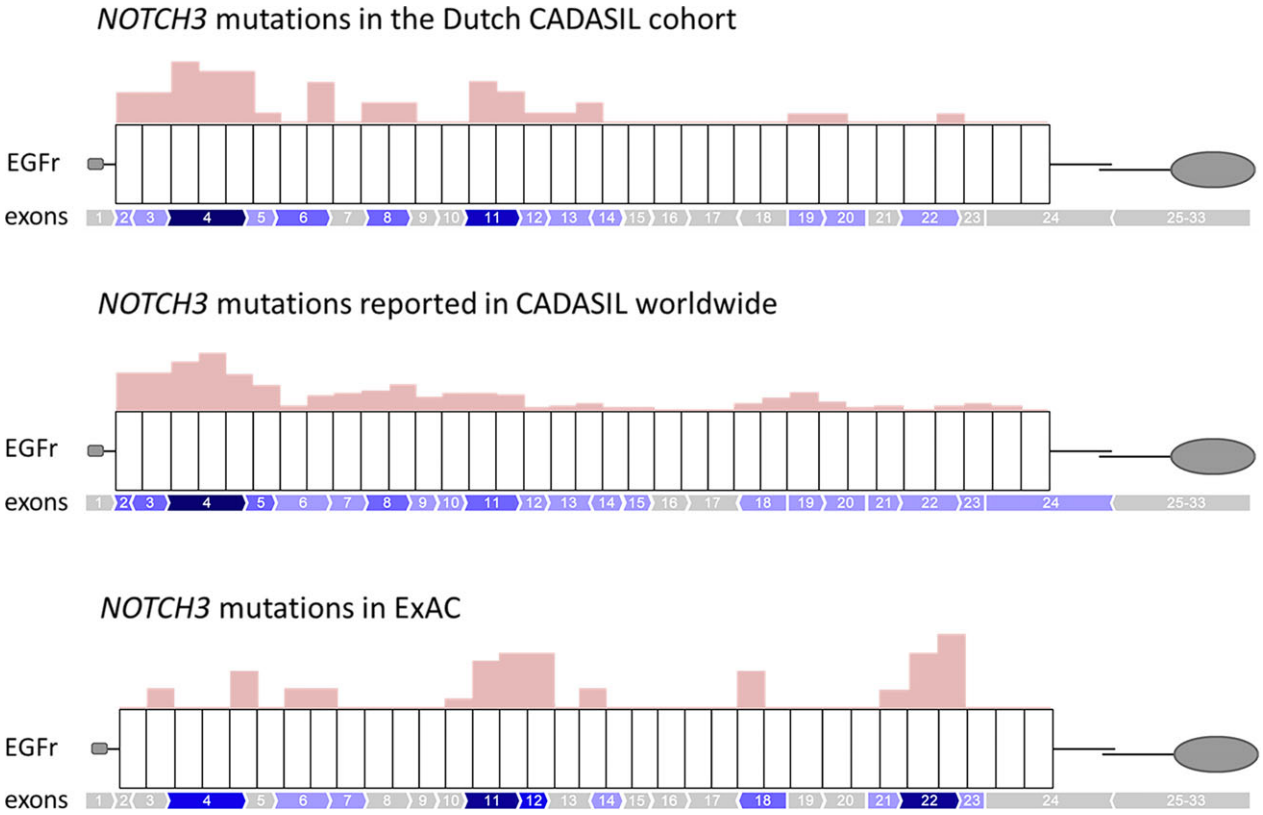

Figure 2. Distribution of distinct EGFr cysteine altering *NOTCH3* mutations in ExAC compared to those reported in cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) patients. Schematic representation of the NOTCH3 protein with 34 epidermal growth factor-like repeat domains (EGFr) and the *NOTCH3* EGFr encoding exons (exon 2–24). Exons 25–33 (and part of exon 24) encode the transmembrane and intracellular domains of NOTCH3; these have never been found to harbor CADASIL-causing mutations. In the Dutch CADASIL registry and in CADASIL patients worldwide, most mutations are located in exon 4, which largely accounts for the predominance of mutations in EGFr domains 1–6. In ExAC, this predominance is not observed, as mutations are most frequent in exons 11–12 and 22, encoding EGFr domains 14–16 and 29–31, respectively. Pink bars above the EGFr domains represent the distribution of mutations across EGFr domains, where the height of the pink bar reflects the percentage of mutations in the respective EGFr domain. Exon colors reflect the number of mutations in each exon: gray indicates exons without mutations; blue indicates exons with mutations. The darker the color blue of the exon, the higher the number of distinct mutations.

patient had a history of hypertension and hyperlipidemia, and one patient had a history of hypercholesterolemia. Their further clinical work-up and medical history was unremarkable. Extensive WMHs consistent with CADASIL were observed on brain imaging in all four patients. None of the patients had lacunes, which are typically seen up to three decades earlier in patients with classical CADASIL (Fig. 4A–C). The individuals who had an MRI scan, made at age 78 and 58 respectively, showed symmetric WMHs in both and microbleeds in one (Fig. 4D–I). These individuals illustrate that *NOTCH3* mutations can be present in clinically asymptomatic or paucisymptomatic individuals over 70 years of age, who for this reason can easily remain undiagnosed with respect to CADASIL.

Discussion

The frequency of EGFr cysteine altering *NOTCH3* mutations in the public exome database ExAC is 3.4/1000. This frequency is 100-fold higher than the current estimated prevalence of CADASIL. Although CADASIL is classically described as a mid-adult onset, severe stroke and dementia syndrome, a later onset and milder disease course is increasingly recognized.[3] Our results suggest that this milder phenotype, which may be indistinguishable from signs and symptoms of sporadic geriatric cerebral small vessel disease, is much more prevalent than recognized to date.

There are four possible explanations for the unexpected high frequency of *NOTCH3* mutations in ExAC, namely

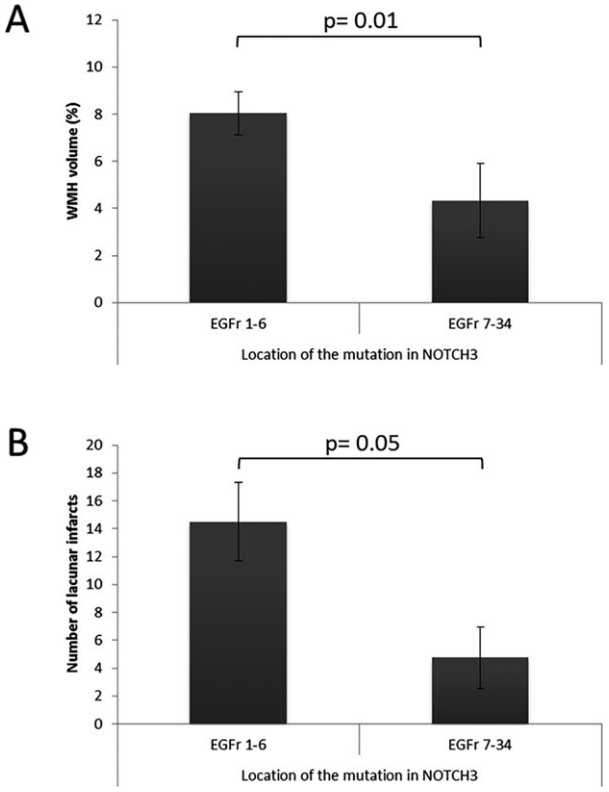

Figure 3. Correlation between EGFr location of the *NOTCH3* mutation and MRI lesion load in cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) patients. MRI quantification data from patients with a mutation in EGFr domains 1–6 ($n = 32$) compared to patients with a mutation in EGFr domains 7–34 ($n = 8$). (A) White matter hyperintensity volume (expressed as the % of total brain parenchymal volume), was significantly higher in patients with mutations in EGFr domains 1–6 than in patients with mutations in EGFr domains 7–34 (8.0% vs. 4.3%, $P = 0.01$). (B) The number of lacunar infarcts showed a similar trend (14.5 vs. 4.8, $P = 0.05$). Bars represent mean ± SEM, corrected for age, sex, smoking and hypertension.

(1) the mutations in ExAC differ from those found in CADASIL and are not pathogenic but are polymorphisms; (2) all the individuals with a *NOTCH3* mutation in ExAC have an undiagnosed or unreported classical CADASIL; (3) some mutations have reduced penetrance in certain contexts, for example, in certain populations; (4) a late-onset, mild CADASIL phenotype which is less readily diagnosed, is much more prevalent than recognized to date.

The limitation of this study is that no phenotypic information is available for individuals in the ExAC database and we therefore cannot definitively confirm or reject any of these possible explanations. However, in our opinion, it is highly unlikely that any of the 25 EGFr cysteine altering *NOTCH3* mutations found in ExAC are polymorphisms, as even the two most frequent mutations in ExAC (p.Arg544Cys and p.Arg1231Cys) have been reported in multiple CADASIL pedigrees.[12,22,23] Furthermore, more than half of the 25 mutations in ExAC have been previously reported in CADASIL pedigrees and all 25 obey the criteria used by DNA-diagnostic laboratories to establish pathogenicity of a *NOTCH3* mutation. It should also be noted that novel EGFr cysteine altering *NOTCH3* mutations are reported in newly diagnosed CADASIL patients at a steady rate, with well over 200 such mutations reported to date.[21]

As ExAC consists of various disease and control cohorts and does not contain any stroke or dementia cohorts, it is unlikely that the 206 individuals with a *NOTCH3* mutation in ExAC all have an undiagnosed classical mid-adult onset CADASIL phenotype. Nonpenetrance in these individuals cannot be excluded but is also unlikely, as this has never been reported for this highly specific type of *NOTCH3* mutation and we did not find nonpenetrance in asymptomatic or paucisymptomatic elderly individuals with a *NOTCH3* mutation in our Dutch registry. Therefore, in our opinion, the most likely explanation for the observed high mutation frequency in ExAC is a much higher prevalence than recognized to date of a very mild, late-onset *NOTCH3* mutation-associated cerebral small vessel disease, which may remain undiagnosed into old age.[24]

Our finding that the EGFr distribution of mutations in ExAC differs from that found in diagnosed CADASIL patients, suggested a possible genotype-phenotype correlation. We hypothesized that mutations located in EGFr domains 1–6 predispose to "classical" CADASIL and, conversely, mutations outside of EGFr domains 1–6 predispose to a much milder phenotype. This would also explain why most CADASIL patients diagnosed to date have a mutation in exon 4, as these have a more severe phenotype and are therefore much more readily diagnosed. We confirmed our hypothesis in a well-characterized independent Dutch CADASIL cohort,[7] in which we found that patients with a mutation in EGFr domains 1–6 have a significantly higher MRI lesion load than those with mutations in EGFr domains 7–34. Interestingly, the paucisymptomatic elderly individuals in our Dutch cohort all had a mutation outside of EGFr domains 1–6, which is also the case in other mildly affected CADASIL patients reported in the literature.[24,25] We can only speculate as to why mutations in certain EGFr domains may have more detrimental effects than others. Possibly, an unpaired cysteine in EGFr domains 1–6, at the N-terminus of the NOTCH3 extracellular domain, more readily interacts with other proteins, leading to a stronger effect on CADASIL-associated protein multimerization.[26,27] Although this novel genotype-phenotype correlation may help predict a predisposition to a milder or a

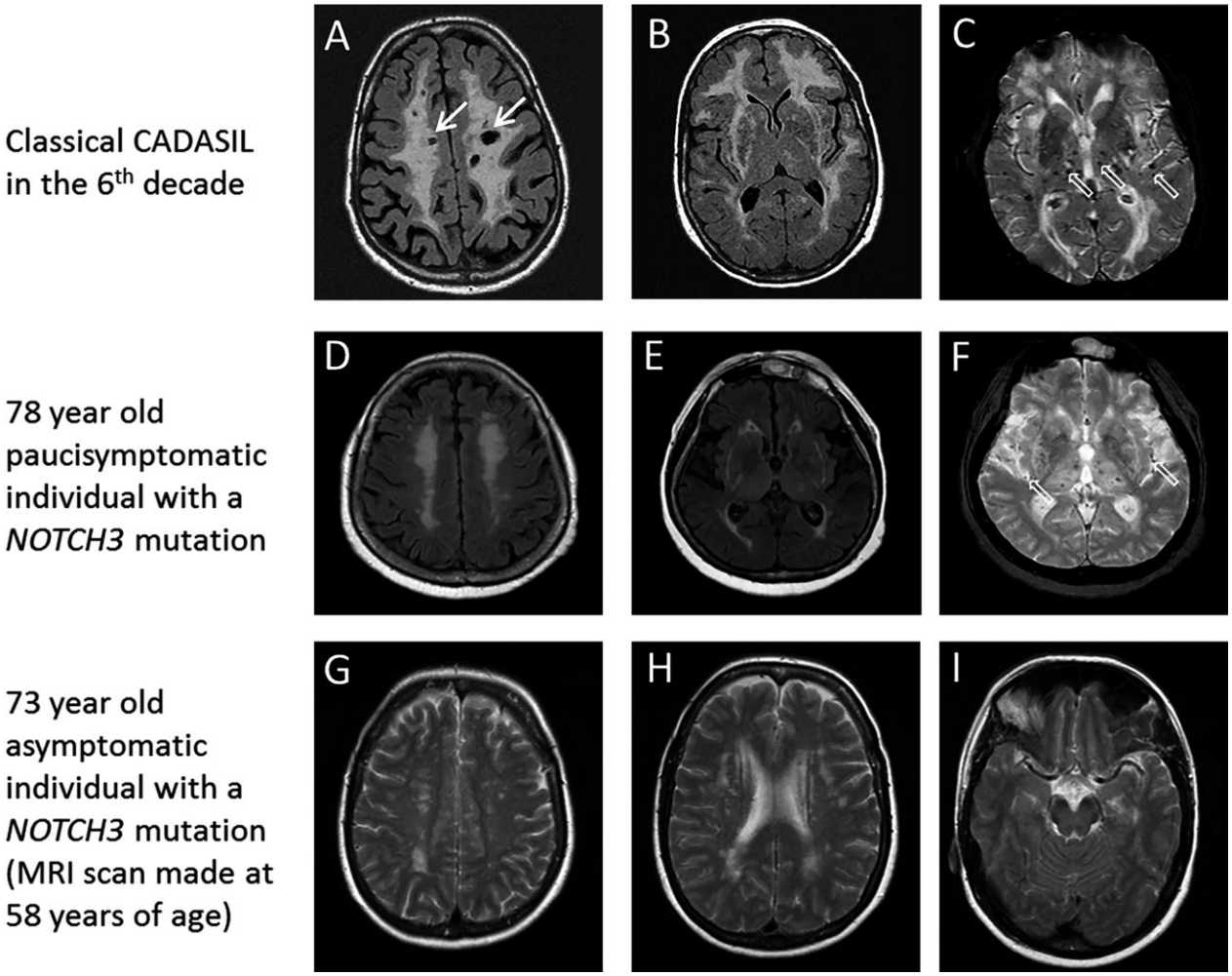

Figure 4. Brain MRI of elderly asymptomatic and paucisymptomatic individuals with a *NOTCH3* mutation compared to MRI in classical cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). (A–C) Brain MRI images of a classical CADASIL phenotype in the 6th decade. (A–B) FLAIR images showing extensive confluent symmetric white matter hyperintensities and multiple lacunar infarcts (arrows indicate random samples). (C) T2*-weighted MRI showing multiple microbleeds (open arrows indicate random samples). (D–F) Brain MRI images of a female, diagnosed with CADASIL after predictive DNA testing at 78 years of age, with only very mild clinical symptoms. (D, E) FLAIR images showing symmetric white matter hyperintensities, but no lacunar infarcts. (F) T2*- weighted MRI showing some small microbleeds (open arrows). (G-I) Dual echo images of a female who was still clinically asymptomatic at 73 years of age. MRI images were made at 58 years of age, showing very mild symmetric white matter hyperintensities, no lacunar infarcts and no microbleeds.

more severe CADASIL phenotype, this needs further delineating in larger cohorts before it can be used in the clinical setting. Also, clearly, factors other than *NOTCH3* genotype play an important role in CADASIL. For example, smoking and hypertension have been shown to be associated with CADASIL disease severity.[13]

A worldwide prevalence of pathogenic *NOTCH3* mutations of 3.4/1000 individuals implies that these mutations may be a major contributor to cerebral small vessel disease in the general population. Indeed, a recent study in patients with adult-onset leukoencephalopathy showed that 21% of these individuals had a pathogenic *NOTCH3* mutation.[28] Similarly, high *NOTCH3* mutation frequencies (18%) were found in a Korean cohort of patients with subcortical vascular cognitive impairment.[29] Conversely, in a cohort of individuals with confluent white matter lesions or lacunes, no cysteine altering *NOTCH3* mutations were found.[30] Future studies in large, well-characterized cohorts are needed to determine the role of cysteine altering *NOTCH3* mutations in the development of white matter hyperintensities and vascular cognitive impairment in the general population.

Our study illustrates how large whole exome datasets can alter mutation-disease paradigms. This was recently

also illustrated by a study reporting an unexpected high frequency of prion disease- associated *PRNP* mutations in ExAC.[31] Although our finding of high *NOTCH3* mutation frequencies shows some similarities to this study, there are also fundamental differences between *NOTCH3* mutations in CADASIL and *PRNP* mutations in prion disease. The most important difference is the fact that many different types of *PRNP* variants have been described, with variable levels of evidence for pathogenicity. This is in sharp contrast to the highly stereotyped nature of pathogenic *NOTCH3* mutations in CADASIL.

In conclusion, we show that the frequency of EGFr cysteine altering *NOTCH3* mutations in a large exome dataset is 3.4/1000 and we discovered the first genotype-phenotype correlation in CADASIL, which is related to the distribution of mutations across *NOTCH3* EGFr domains. These findings fundamentally change the CADASIL disease paradigm and cause us to redefine the phenotypic spectrum associated with EGFr cysteine altering *NOTCH3* mutations.

Acknowledgments

The authors thank Michael Liem for quantification of MRI data, and the Leiden Laboratory for Diagnostic Genome Analysis (LDGA) for *NOTCH3* mutation analysis of Dutch CADASIL patients. This work was financially supported by ZonMw, The Netherlands Organization for Health Research and Development, grant no. 40-41900-98-018.

Author contributions

JR and SLO designed the study. JR, HD, GG, MBQ, JP, JvdG, and SLO analyzed the data. MB supervised DNA analysis. JR and SLO wrote the manuscript. All authors provided scientific input and constructive criticism on the manuscript, and approved the final version of the report for publication.

Conflicts of Interest

NOTCH3 antisense therapies have been patented by the Leiden University Medical Center. As co-inventors on this patent HD and SLO are entitled to a share of potential royalties.

References

1. Joutel A, Corpechot C, Ducros A, et al. Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. Nature 1996;383:707–710.
2. Chabriat H, Joutel A, Dichgans M, et al. Cadasil. Lancet Neurol 2009;8:643–653.
3. Moreton FC, Razvi SS, Davidson R, et al. Changing clinical patterns and increasing prevalence in CADASIL. Acta Neurol Scand 2014;130:197–203.
4. Liem MK, Lesnik Oberstein SA, Haan J, et al. MRI correlates of cognitive decline in CADASIL: a 7-year follow-up study. Neurology 2009;72:143–148.
5. Viswanathan A, Gschwendtner A, Guichard JP, et al. Lacunar lesions are independently associated with disability and cognitive impairment in CADASIL. Neurology 2007;69:172–179.
6. Peters N, Holtmannspotter M, Opherk C, et al. Brain volume changes in CADASIL: a serial MRI study in pure subcortical ischemic vascular disease. Neurology 2006;66:1517–1522.
7. van den Boom R, Lesnik Oberstein SA, Ferrari MD, et al. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy: MR imaging findings at different ages–3rd-6th decades. Radiology 2003;229:683–690.
8. Lesnik Oberstein SA, van den Boom R, Middelkoop HA, et al. Incipient CADASIL. Arch Neurol 2003;60:707–712.
9. Chabriat H, Levy C, Taillia H, et al. Patterns of MRI lesions in CADASIL. Neurology 1998;51:452–457.
10. Lesnik Oberstein SA, vanden Boom R, vanBuchem MA. Cerebral microbleeds in CADASIL. Neurology 2001; 57:1066–1070.
11. Kim YE, Yoon CW, Seo SW, et al. Spectrum of NOTCH3 mutations in Korean patients with clinically suspicious cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy. Neurobiol Aging 2014;35:726.
12. Joutel A, Vahedi K, Corpechot C, et al. Strong clustering and stereotyped nature of Notch3 mutations in CADASIL patients. Lancet 1997;350:1511–1515.
13. Adib-Samii P, Brice G, Martin RJ, et al. Clinical spectrum of CADASIL and the effect of cardiovascular risk factors on phenotype: study in 200 consecutively recruited individuals. Stroke 2010;41:630–634.
14. Opherk C, Peters N, Herzog J, et al. Long-term prognosis and causes of death in CADASIL: a retrospective study in 411 patients. Brain 2004;127:2533–2539.
15. Bianchi S, Zicari E, Carluccio A, et al. CADASIL in central Italy: a retrospective clinical and genetic study in 229 patients. J Neurol 2015;262:134–141.
16. Ampuero I, Alegre-Abarrategui J, Rodal I, et al. On the diagnosis of CADASIL. J Alzheimers Dis 2009;17: 787–794.
17. Yin X, Wu D, Wan J, et al. Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy: Phenotypic and mutational spectrum in patients from mainland China. Int J Neurosci 2015;125:585–592.
18. Narayan SK, Gorman G, Kalaria RN, et al. The minimum prevalence of CADASIL in northeast England. Neurology 2012;78:1025–1027.

19. Razvi SS, Davidson R, Bone I, et al. The prevalence of cerebral autosomal dominant arteriopathy with subcortical infarcts and leucoencephalopathy (CADASIL) in the west of Scotland. J Neurol Neurosurg Psychiatry 2005;76:739–741.
20. Duering M, Karpinska A, Rosner S, et al. Co-aggregate formation of CADASIL-mutant NOTCH3: a single-particle analysis. Hum Mol Genet 2011;20:3256–3265.
21. Rutten JW, Haan J, Terwindt GM, et al. Interpretation of NOTCH3 mutations in the diagnosis of CADASIL. Expert Rev Mol Diagn 2014;14:593–603.
22. Lee YC, Liu CS, Chang MH, et al. Population-specific spectrum of NOTCH3 mutations, MRI features and founder effect of CADASIL in Chinese. J Neurol 2009;256:249–255.
23. Abou Al-Shaar H, Qadi N, Al-Hamed MH, et al. Phenotypic comparison of individuals with homozygous or heterozygous mutation of NOTCH3 in a large CADASIL family. J Neurol Sci 2016;367:239–243.
24. Pescini F, Bianchi S, Salvadori E, et al. A pathogenic mutation on exon 21 of the NOTCH3 gene causing CADASIL in an octogenarian paucisymptomatic patient. J Neurol Sci 2008;267:170–173.
25. Lee YC, Yang AH, Soong BW. The remarkably variable expressivity of CADASIL: report of a minimally symptomatic man at an advanced age. J Neurol 2009;256:1026–1027.
26. Monet-Lepretre M, Haddad I, Baron-Menguy C, et al. Abnormal recruitment of extracellular matrix proteins by excess Notch3ECD: a new pathomechanism in CADASIL. Brain 2013;136:1830–1845.
27. Kast J, Hanecker P, Beaufort N, et al. Sequestration of latent TGF-beta binding protein 1 into CADASIL-related Notch3-ECD deposits. Acta Neuropathol Commun 2014;2:96.
28. Ayrignac X, Carra-Dalliere C, Menjot de CN, et al. Adult-onset genetic leukoencephalopathies: a MRI pattern-based approach in a comprehensive study of 154 patients. Brain 2015;138:284–292.
29. Yoon CW, Kim YE, Seo SW, et al. NOTCH3 variants in patients with subcortical vascular cognitive impairment: a comparison with typical CADASIL patients. Neurobiol Aging 2015;36:2443–2447.
30. Schmidt H, Zeginigg M, Wiltgen M, et al. Genetic variants of the NOTCH3 gene in the elderly and magnetic resonance imaging correlates of age-related cerebral small vessel disease. Brain 2011;134:3384–3397.
31. Minikel EV, Vallabh SM, Lek M, et al. Quantifying prion disease penetrance using large population control cohorts. Sci Transl Med 2016;8:322ra9.
32. Dichgans M, Filippi M, Bruning R, et al. Quantitative MRI in CADASIL: correlation with disability and cognitive prformance. Neurology 1999;52:1361–1367.
33. Oberstein SA, Ferrari MD, Bakker E, et al. Diagnostic Notch3 sequence analysis in CADASIL: three new mutations in Dutch patients. Dutch CADASIL Research Group. Neurology 1999;52:1913–1915.
34. Bentley P, Wang T, Malik O, et al. CADASIL with cord involvement associated with a novel and atypical NOTCH3 mutation. J Neurol Neurosurg Psychiatry 2011;82:855–860.
35. Kim Y, Kim JS, Kim G, et al. Two novel mutations of the NOTCH3 gene in Korean patients with CADASIL. Mutat Res 2006;593:116–120.
36. Escary JL, Cecillon M, Maciazek J, et al. Evaluation of DHPLC analysis in mutational scanning of Notch3, a gene with a high G-C content. Hum Mutat 2000;16:518–526.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Pathogenic *NOTCH3* mutations in the Netherlands. Between 1998 and 2014, 45 distinct mutations have been detected in the Netherlands. Six mutations are especially prevalent (shown in bold), these include both (Dutch) founder mutations and recurrent mutations (Unpublished). *mutations which have not been previously reported, †mutation reported two times in a homozygous state.

