

Multi-infarct dementia of Swedish type is caused by a 3'UTR mutation of *COL4A1*

M. Siitonen^{1*}, A. Börjesson-Hanson^{2*}, M. Pöyhönen^{3,4}, A. Ora⁵, P. Pasanen^{1,6}, J. Bras^{7,8}, S. Kern², J. Kern², O. Andersen⁹, H. Stanescu¹⁰, R. Kleta¹⁰, M. Baumann¹¹, R. Kalaria¹², H. Kalimo¹³, A. Singleton¹⁴, J. Hardy^{7,15}, M. Viitanen^{16,17}, L. Myllykangas^{13#} and R. Guerreiro^{7,8#}

1) Department of Medical Biochemistry and Genetics, University of Turku, Turku, Finland

2) Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden

3) Department of Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland

4) Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland

5) Department of Applied Physics, Aalto University, Espoo, Finland

6) Tyks Microbiology and Genetics, Turku University Hospital, Turku, Finland

7) Department of Molecular Neuroscience, Institute of Neurology, University College London, London, UK

8) Department of Medical Sciences and Institute of Biomedicine - iBiMED, University of Aveiro, Aveiro, Portugal

9) Department of Clinical Neuroscience, Institute of Neuroscience and Physiology, Sahlgrenska Academy at the University of Gothenburg, Gothenburg, Sweden

10) Centre for Nephrology University College London London UK.

11) Meilahti Clinical Proteomics Core Unit, Department of Biochemistry and Developmental Biology, Biomedicum Helsinki, University of Helsinki, Helsinki, Finland

12) Institute of Neuroscience, Newcastle University, Campus for Ageing and Vitality, Newcastle upon Tyne, UK

13) Department of Pathology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

14) Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA.

15) Reta Lila Weston Institute, UCL Institute of Neurology, London, UK

16) Department of Neurobiology, Care Sciences and Society, Karolinska Institutet and Karolinska University Hospital Huddinge, Stockholm, Sweden

17) Department of Geriatrics, University of Turku and Turku City Hospital, Turku, Finland.

** These authors contributed equally to this study*

These authors contributed equally to this study

Corresponding author:

Rita Guerreiro

Department of Molecular Neuroscience, Institute of Neurology, University College London

1 Wakefield Street, London WC1N 1PJ, England

Telephone: +44 (0) 207 679 4256

Email: r.guerreiro@ucl.ac.uk

Running title: 3'UTR mutation of *COL4A1* causes hMID in a Swedish family

Cerebral small vessel diseases (cSVD) often present as sporadic conditions but several monogenic families have also been reported (Hagel *et al.*, 2004; Herve *et al.*, 2012). In 1977, Sourander and Wålinder described a family with an autosomal dominant cerebrovascular disease manifesting with transient ischaemic attacks/strokes, neuropsychiatric symptoms and progressive cognitive decline. Thirty years later it was proved that this family did not have mutations in *NOTCH3*, excluding the initially suspected diagnosis of cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). Consequently it was concluded that this family presented a new cSVD, which was named hereditary multi-infarct dementia (hMID) of the Swedish type (Low *et al.*, 2007).

In order to identify the genetic cause of disease in this Swedish hMID family we performed whole-exome sequencing (WES) and genetic linkage analysis. Twenty-one family members participated in this study: 10 were affected, 10 were unaffected and one participant was an unrelated spouse used as a control (Figure 1A).

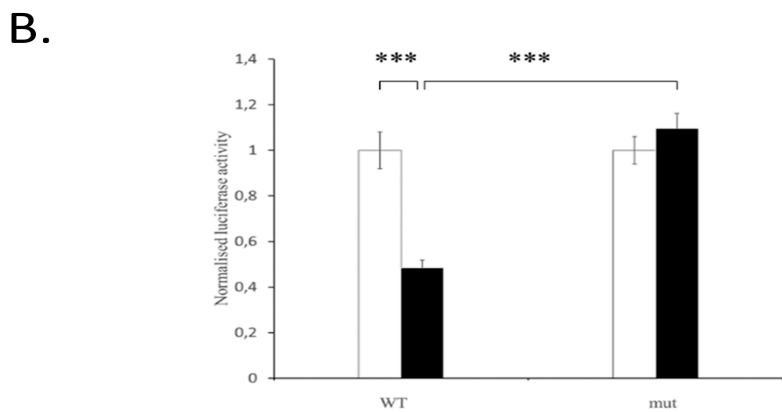
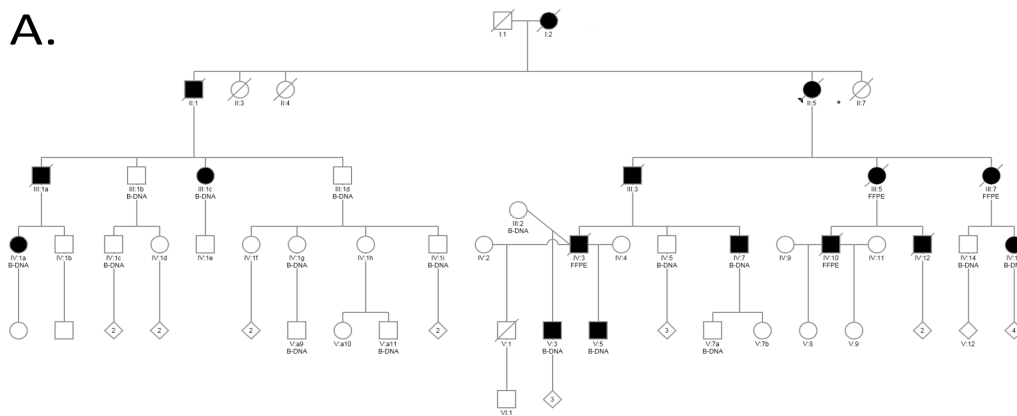


Figure 1: A. Pedigree of the Swedish hMID family. Family members included in the study are marked with B-DNA (blood derived DNA) or FFPE (formalin fixed paraffin embedded

tissue) according to the sample type available. Black symbols represent affected individuals and white symbols represent healthy family members. Diagonal lines indicate deceased individuals. The arrowhead indicates the proband (Individual II:5). B. Luciferase assay. HEK-293T cells were transfected with pMIR-REPORT luciferase wild-type (WT) or mutant (mut). Cells were co-transfected either with miR-29b-3p (black bars) or negative control microRNA (white bars). Normalized luciferase activity of cells transfected with the WT construct was significantly decreased by -miR-29b-3p, compared to cells transfected with a mimic-negative control. Luciferase activity was not altered in cells transfected with a mutated construct ($***p < 0.001$, 2-sided Student *t* test). Error bars indicate mean standard deviation.

Blood samples (n=17) were collected and DNA was extracted by standard methods after written informed consent was provided by all family members taking part in this study. For 4 patients only formalin fixed paraffin embedded (FFPE) tissues were available.

Four patients (III:1c, IV:7, IV:16 and V:3) and 2 unaffected family members (IV:5 and IV:14) were selected for WES. Exomes were prepared using the SeqCap EZ Human Exome Library version 2.0 (Roche Nimblegen Inc) and sequencing runs were performed on HiSeq 2000 (Illumina). Sequencing reads were aligned to GRCh37/hg19 using BWA (Li and Durbin, 2010) and variants were called according to GATK's standard best practices v3 (McKenna *et al.*, 2010; DePristo *et al.*, 2011). Following variant calling, annotation was performed using SnpEff (Cingolani *et al.*, 2012). For the linkage study, whole genome genotyping was performed for 12 blood-derived DNA samples using HumanOmniExpress Bead chips (Illumina). Parametric multipoint linkage analysis was performed using Allegro (Gudbjartsson *et al.*, 2000) under a fully penetrant autosomal dominant model.

Data analyses were based on an autosomal dominant mode of inheritance of the disease and the hypothesis that the underlying mutation was not present in neurologically healthy control individuals or in the general population (**Table 1**). Validation of variants found by WES was done using Sanger sequencing with BigDye Terminator version 3.1 chemistry (Applied Biosystems).

Parametric multipoint linkage analysis identified four peaks on chromosomes 10, 11, 12 and 13 achieving LOD scores > 2 . When these regions were compared to the WES data we identified three variants present in affected family members and absent in healthy relatives: one in *SPOCK2* and two variants in *COL4A1* (**Table 1**).

Table 1: Linkage regions with $LOD > 2$ and the variants identified by WES located in these chromosomal areas.

Chr	Chr location	dbSNP	LS	Variants	Gmaf
10	7349891-76372030	rs10823837- rs4746209	2,352	<i>SPOCK2</i> c.*11G>A	0,379
11	47929846- 49000550	rs6485795- rs11040198	2,294	No variants identified	
12	85165879- 87281210	rs11116595- rs7316774	2,348	No variants identified	
13	109327788- 111067000	rs9284246- rs10851243	2,407	<i>COL4A1</i> c.*32G>A <i>COL4A1</i> n.249C>T, c.4470C>T	- 0,376

Chr: Chromosome; *Chr location:* Chromosomal location (hg19); *dbSNP:* dbSNP accession numbers; *LS:* Logarithm of odds score; *Variants:* variants identified by WES in the region; *Gmaf:* global minor allele frequency in the Exome Aggregation Consortium (ExAC). Minor allele frequencies of 0,4 in the general population for both *SPOCK2:c.*11G>A* and *COL4A1:c.4470C>T* were considered to be too high for a mutation causative of a rare disease as Swedish hMID.

Only one of these variants (*COL4A1:c.*32G>A*) was found to segregate with the disease in the extended family and was absent from population databases. Although the variant is located in the 3'UTR of *COL4A1* both gnomAD and ExAC databases report variants in this locus in a minimum of 117,613 and 44,384 individuals, respectively (Lek *et al.*, 2016). The segregation of *COL4A1 c.*32G>A* with the disease was confirmed using Sanger sequencing. All affected cases had the variant and none of the older unaffected cases (age > 40 years) carried it. One younger, currently unaffected, family member also carried the *COL4A1 c.*32G>A*.

The c.*32G>A mutation is located in the 3' UTR region of *COL4A1*, and may affect the binding site of miR-29, located in this region. To test this hypothesis we performed a microRNA transfection study combined with luciferase reporter assay. 197 bp fragment of wild-type and mutated target site were amplified by PCR from patients' genomic DNA. The inserts were validated by sequencing. Amplified target region was digested with Hind III/Spe1, cloned into the pMIR REPORT Luciferase plasmid (AppliedBiosystems).

HEK293T cells (DMEM, 10% FCS serum with penicillin (100 U/ml) and streptomycin (100 µg/ml) in humidified air containing 5% CO₂ at 37°C) were plated in 24-well plates. At 80% confluence, 100 ng of empty, wild-type, or mutated plasmids were cotransfected with 25 pmol of either miRIDIAN hsa-miR-29b-3p or negative control (Dharmacon), using DharmaFECT Duo 2.5 µl in each well (Dharmacon). The triplicate samples were lysed with 1% NP40, 150 mM NaCl and 25 mM Tris, pH 7.6, and firefly luciferase activities were measured 36 hours after transfection using the 1000 Assay System (Promega) and analyzed with BioTek, Cytostation 5. The results suggested that the *COL4A1* c.*32G>A mutation affects miR-29 binding and hence leads to upregulation of *COL4A1* (**Figure 1B**).

COL4A1 mutations have been reported as the cause of a wide variety of autosomal dominant diseases being associated with variable phenotypes (Lemmens *et al.*, 2013). These include: porencephaly 1 (OMIM #175780); small vessel disease of the brain with or without ocular anomalies (BSVD, OMIM #607595); retinal arterial tortuosity (RATOR, OMIM #180000); hereditary angiopathy with nephropathy, aneurysms and muscle cramp (HANAC, OMIM #611773); Walker-Warburg syndrome (Labelle-Dumais *et al.*, 2011) and pontine autosomal dominant microangiopathy with leukoencephalopathy (PADMAL) (Verdura *et al.*, 2016).

Swedish hMID is a cerebral small vessel disease characterized by multifocal impaired cerebral blood flow resulting in multiple infarctions. Clinically and pathologically it fits within the expanding phenotypic group of *COL4A1* related disorders, most closely resembling PADMAL with lacunar infarcts in the subcortical and pontine areas (Sourander and Walinder, 1977; Hagel *et al.*, 2004; Verdura *et al.*, 2016).

The 3'UTR *COL4A1* variant identified here presented complete segregation with the disease in this family, being identified in all affected and absent in all older (>40 years of age) unaffected family members. One younger, healthy subject (V:7a) carried the *COL4A1* c.*32G>A variant, suggesting the possibility of currently being in an asymptomatic stage of the disease.

A recent publication by Verdura and colleagues identified mutations in *COL4A1* 3'UTR as the cause of cSVD in six families, including PADMAL cases. The mutations identified also affected the binding site for miR-29 micro-RNA located within the 3'UTR of *COL4A1*, and were shown to lead to upregulation of *COL4A1* mRNA expression (Verdura *et al.*, 2016). Although the variant found in this Swedish hMID family is novel, it disrupts the same miR-29 binding site, adding support to the pathogenicity of the mutation and suggesting that *COL4A1* upregulation is a central pathogenic mechanism both in Swedish hMID and PADMAL. The similarities at clinical and pathological levels also support this view: both

diseases are characterized by fibrohyalinosis and elastosis of small arterioles with atrophy of media and proliferation of the intima. These changes result in multiple lacunar infarcts in the basal ganglia, thalamus, periventricular white matter and pons, and in cortical and white matter atrophy. At the EM level, thickening of the basement membrane is observed in both diseases (Sourander and Walinder, 1977; Hagel *et al.*, 2004; Low *et al.*, 2007; Verdura *et al.*, 2016), and the clinical findings include cognitive impairment and progressive dementia, strokes, as well as mood and gait disturbances (Hagel *et al.*, 2004; Low *et al.*, 2007; Craggs *et al.*, 2013). In our previous study, immunostaining of COL4 revealed increased levels of staining in walls of small cerebral arteries both in PADMAL and Swedish hMID cases. However, the investigation of the sclerotic index showed some regional differences between the diseases: PADMAL seemed to affect the vessels of the frontal region more than those of the basal ganglia, whereas hMID cases showed the opposite effect (Craggs *et al.*, 2013). Furthermore, no haemorrhages have been described in subjects with PADMAL, while one hMID subject with anticoagulative treatment was reported to suffer from a massive haemorrhage (Sourander and Walinder, 1977).

As previously proposed, perturbations of the cerebrovascular matrisome (the group of proteins both constituting and associated with the extracellular matrix) can represent a convergent pathologic pathway in monogenic small vessel diseases (Joutel *et al.*, 2016). Still, further studies are needed to clarify the detailed pathogenic molecular mechanisms behind these diseases and to understand the phenotypic differences arising from mutations in the same micro-RNA binding site.

Legends for figures

Figure 1: A. Pedigree of the Swedish hMID family. Family members included in the study are marked with B-DNA (blood derived DNA) or FFPE (formalin fixed paraffin embedded tissue) according to the sample type available. Black symbols represent affected individuals and white symbols represent healthy family members. Diagonal lines indicate deceased individuals. The arrowhead indicates the proband (Individual II:5). B. Luciferase assay. HEK-293T cells were transfected with pMIR-REPORT luciferase wild-type (WT) or mutant (mut). Cells were co-transfected either with miR-29b-3p (black bars) or negative control microRNA (white bars). Normalized luciferase activity of cells transfected with the WT construct was significantly decreased by -miR-29b-3p, compared to cells transfected with a

mimic-negative control. Luciferase activity was not altered in cells transfected with a mutated construct (**p < 0.001, 2-sided Student t test). Error bars indicate mean standard deviation.

Acknowledgements

We are very grateful to all the family members that participated in this study. The authors acknowledge Leena Saikko for technical help and the Päivikki and Sakari Sohlberg Foundation. The authors would like to thank the Genome Aggregation Database (gnomAD) and the groups that provided exome and genome variant data to this resource. A full list of contributing groups can be found at <http://gnomad.broadinstitute.org/about>.

Conflict of Interest and Sources of Funding

The authors have no conflicts of interest. This work was funded by the Sigrid Juselius Foundation, Academy of Finland project numbers 115906 and 294817; the EVO Research Funds of the University Hospitals of Helsinki and Turku; the City Hospital of Turku and Finnish Cultural Foundation, Varsinais-Suomi regional fund; research funds from Sahlgrenska University Hospital, Gothenburg, Sweden; Alzheimer's Research UK and by research fellowships awarded by Alzheimer's Society to Jose Bras and Rita Guerreiro.

References

- Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, *et al.* 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* 2012; 6(2): 80-92.
- Craggs LJ, Hagel C, Kuhlenbaeumer G, Borjesson-Hanson A, Andersen O, Viitanen M, *et al.* Quantitative vascular pathology and phenotyping familial and sporadic cerebral small vessel diseases. *Brain pathology* 2013; 23(5): 547-57.
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; 43(5): 491-8.
- Gudbjartsson DF, Jonasson K, Frigge ML, Kong A. Allegro, a new computer program for multipoint linkage analysis. *Nat Genet* 2000; 25(1): 12-3.
- Hagel C, Groden C, Niemeyer R, Stavrou D, Colmant HJ. Subcortical angiopathic encephalopathy in a German kindred suggests an autosomal dominant disorder distinct from CADASIL. *Acta neuropathologica* 2004; 108(3): 231-40.
- Herve D, Chabriat H, Rigal M, Dalloz MA, Kawkabani Marchini A, De Lepeleire J, *et al.* A novel hereditary extensive vascular leukoencephalopathy mapping to chromosome 20q13. *Neurology* 2012; 79(23): 2283-7.
- Joutel A, Haddad I, Ratelade J, Nelson MT. Perturbations of the cerebrovascular matrisome: A convergent mechanism in small vessel disease of the brain? *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism* 2016; 36(1): 143-57.
- Labelle-Dumais C, Dilworth DJ, Harrington EP, de Leau M, Lyons D, Kabaeva Z, *et al.* COL4A1 mutations cause ocular dysgenesis, neuronal localization defects, and myopathy in mice and Walker-Warburg syndrome in humans. *PLoS Genet* 2011; 7(5): e1002062.
- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; 536(7616): 285-91.
- Lemmens R, Maugeri A, Niessen HW, Goris A, Tousseyn T, Demaerel P, *et al.* Novel COL4A1 mutations cause cerebral small vessel disease by haploinsufficiency. *Hum Mol Genet* 2013; 22(2): 391-7.
- Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010; 26(5): 589-95.

Low WC, Junna M, Borjesson-Hanson A, Morris CM, Moss TH, Stevens DL, *et al.* Hereditary multi-infarct dementia of the Swedish type is a novel disorder different from NOTCH3 causing CADASIL. *Brain* 2007; 130(Pt 2): 357-67.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research* 2010; 20(9): 1297-303.

Sourander P, Walinder J. Hereditary multi-infarct dementia. Morphological and clinical studies of a new disease. *Acta neuropathologica* 1977; 39(3): 247-54.

Verdura E, Herve D, Bergametti F, Jacquet C, Morvan T, Prieto-Morin C, *et al.* Disruption of a miR-29 binding site leading to COL4A1 upregulation causes pontine autosomal dominant microangiopathy with leukoencephalopathy. *Ann Neurol* 2016; 80(5): 741-53.