A rare missense variant in RCL1 segregates with depression in extended families

Najaf Amin1,10,* (PhD), Femke M.S. de Vrij2,10 (PhD), Maarouf Baghdadi2,10 (MSc), Rutger W. W. Brouwer3 (PhD), Jeroen J.G. van Rooij4 (MSc), Olivera Jovanonva5 (MSc), André G. Uitterlinden4,5 (PhD), Albert Hofman5,6 (PhD), Harry L.A. Janssen5,7,8 (PhD), Sarwa Darwish Murad5,7 (PhD), Robert Kraaij4,5 (MSc), Jeffrey Stedehouder2 (MSc), Mirjam C.G.N. van den Hout3 (PhD), Johan M. Kros9 (PhD), Wilfred F.J. van Ijcken3 (PhD), Henning Tiemeier5 (PhD), Steven A. Kushner2,11 (PhD), Cornelia M. van Duijn1,11 (PhD)

1. Genetic Epidemiology Unit, Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands
2. Department of Psychiatry, Erasmus MC, Rotterdam, the Netherlands
3. Department of Cell Biology, Center for Biomics, Erasmus MC, Rotterdam, the Netherlands
4. Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands
5. Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands
6. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
7. Department of gastroenterology and hepatology, Erasmus MC, Rotterdam, the Netherlands
8. Department of Hepatology, University Health Network Toronto Western & General Hospital
9. Department of Pathology, Erasmus MC, Rotterdam, the Netherlands
10. These authors contributed equally to this work
11. These authors contributed equally to this work

* Correspondence to:
Dr. Najaf Amin (n.amin@erasmusmc.nl)
Phone: 0031 (0)10 7038125
Department of Epidemiology, Erasmus University Medical Center,
P.O. Box 2040, 3000CA, Rotterdam, The Netherlands
Exome sequencing/genotyping

Exomes of 1,336 individuals from the ERF study were sequenced at the Center for Biomics (Department of Cell Biology, Erasmus MC, Rotterdam, The Netherlands). Sequencing was performed at a median depth of 57x (mean = 74x, Agilent, version 4 capture) on an Illumina Hiseq2000 sequencer using the TruSeq Version 3 protocol. The sequence reads were aligned to the human genome build 19 (hg19) using Burrows-Wheeler Aligner (BWA) and the NARWHAL pipeline (1, 2). Aligned reads were processed using the IndelRealigner, MarkDuplicates, and Table Recalibration tools from the Genome Analysis Toolkit (GATK) (3) and Picard (http://picard.sourceforge.net) to remove systematic biases and recalibrate the PHRED quality scores. After processing, variants were called using the Unified Genotyper tool from the GATK v2.3(3). For each sample, at least 4 gigabases of sequence was aligned to the genome. Functional annotations were performed using SeattleSeq (http://snp.gs.washington.edu/SeattleSeqAnnotation131). A total of approximately 1.4 million Single Nucleotide Variants (SNVs) were called. After removing variants with low quality (QUAL < 150), diverging from HWE (p-value < 10^-6), or having a low call rate (< 99%), and samples with a low call rate (< 90%) and ethnic outliers identified using a principal component analysis (PCA) with 1000 Genomes data (Figure S16), we retrieved 543,954 high-quality SNVs in 1,327 individuals. Of these, 1,247 individuals (60% women, mean age = 48.5 years) were assessed for depressive symptoms.

Participants from ERF, whose exomes were not sequenced (N = 1,527) were genotyped on the Illumina Infinium HumanExome BeadChip, version 1.1. Calling was performed with GenomeStudio. We removed subjects with a call rate < 0.95, IBS > 0.99 and heterozygote ratio > 0.60. Ethnic outliers identified using a principal component analysis (1)
with 1000 Genomes data (Figure S16) and individuals with sex discrepancies were also removed. The SNVs that were monomorphic in our sample or had a call rate < 0.95 were removed. After quality control we retrieved about 70,000 polymorphic SNVs in 1,515 subjects. Of these, 840 individuals (54% women, mean age = 50 years) were assessed for depressive symptoms.

In RS, exomes of 2628 randomly selected individuals from the RS-I population were sequenced at the Human Genotyping facility of the Erasmus MC (Department of Internal Medicine). Sequencing was performed at an average depth of 54x (Nimblegen SeqCap EZ V2 capture) on an Illumina Hiseq2000 sequencer using the TrueSeq Version 3 protocol. Sequence reads were aligned to hg19 using Burrows-Wheeler Aligner (BWA) (2). Aligned reads were processed using Picard’s MarkDuplicates, SAMtools (4), and the Indel Realignment and Base Quality Score Recalibration tools from Genome Analysis Toolkit (GATK) (3). Genetic variants were called using the HaplotypeCaller from GATK. Samples with low concordance to array genotyping (< 95%) low call rate (< 90%), or that differed ≥4 standard deviations on either the number of detected variants per sample, transition to transversion ratio, or heterozygote to homozygote ratios and were removed. Single Nucleotide Variants (SNVs) with a low call rate (< 90%) or diverged from Hardy-Weinberg equilibrium (HWE; p-value < 10^{-8}) were also removed. Ethnic outliers identified using a principal component analysis with 1000 Genomes data (Figure S16) were removed. The final dataset consisted of 600,806 SNVs among 2356 individuals, of which 1265 were assessed for depressive symptoms (CES-D scale). The average age at examination was 72.3 years and the majority were women (59%; Table S1). File handling and formatting was
implemented using VCFtools(5) and PLINK(6). Variant annotation was performed using SeattleSeq (http://snp.gs.washington.edu/SeattleSeqAnnotation138/).

References

1. Brouwer RW, van den Hout MC, Grosveld FG, van Ijcken WF: NARWHAL, a primary analysis pipeline for NGS data. Bioinformatics 2012; 28: 284-5
2. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25: 1754-60
3. 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 Res 2010; 20: 1297-303
4. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al: The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009; 25: 2078-9
5. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al: The variant call format and VCFtools. Bioinformatics 2011; 27: 2156-8
6. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al: PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559-75

Supplementary Table 1. Demographic characteristics of the studied populations.

|                | ERF (discovery) | RS (replication) |
|----------------|-----------------|------------------|
|                | Linkage (N=218) | Association (N=2393) | N=1604 |
| age (range)    | 58 (19-89)      | 50.5 (18-90)      | 72.3 (61-98) |
| %women         | 63              | 58                | 58 |
| HADS-D         | 14.8            | 6                 | NA |
| CES-D          | 27.4            | 10.6              | 4.8 |

Supplementary Table 2. Details of human donors of fresh frozen middle frontal gyrus brain tissue.

| Brain Number | Age | Sex | Cause of death                  | Brain Weight |
|--------------|-----|-----|---------------------------------|--------------|
| Donor 1      | 79  | M   | Pulmonary dysfunction           | 1482         |
| Donor 2      | 61  | F   | Pulmonary dysfunction           | 1232         |
| Donor 3      | 81  | F   | Intraparenchymal hemorrhage     | 1250         |
Supplementary Table 3. All missense variants shared by two or more affected haplotype carriers in family 1 in the 9p22-24 region.

| SNP       | 38756 | 38753 | 38082 | 16413 | 38081 | 33522 | 15469 | 16411 | MAF | position | chrom | function | polyPhen | consScoreGERP | gene | count |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-----|----------|-------|-----------|----------|--------------|------|-------|
| rs6415788 | 0     | 0     | 1     | 1     | 1     | 1     | 1     | 0     | 0.3455 | 4118111  | 9     | missense  | 0        | 5.37         | GLIS3 | 5      |
| rs41302077 | 2     | 2     | 1     | 0     | 1     | 0     | 0     | 1     | 0.05727 | 4617999  | 9     | missense  | 0.989    | 2.02         | SPATA6L| 5      |
| rs45463394 | 2     | 2     | 1     | 0     | 1     | 0     | 0     | 1     | 0.05727 | 4625415  | 9     | missense  | 0        | 2.3          | SPATA6L| 5      |
| rs34250374 | 0     | 0     | 0     | 1     | 0     | 0     | 0     | 1     | 0.09766 | 4662394  | 9     | missense  | 0        | 3.55         | PPAPDC2| 2      |
| rs34250374 | 0     | 0     | 0     | 1     | 0     | 0     | 0     | 1     | 0.09766 | 4662394  | 9     | missense  | 0        | 3.55         | SPATA6L| 2      |
| rs115482041| 2     | 1     | 1     | 1     | 1     | 0     | 0     | 1     | 0.01319 | 4860267  | 9     | missense  | 0.684    | 4.04         | RCL1  | 6      |
| rs29296067 | 0     | 0     | 0     | 0     | 1     | 2     | 0     | 0     | 0.2704  | 6984236  | 9     | missense  | 0.014    | -0.93        | KDM4C | 2      |
| rs35826653 | 0     | 0     | 1     | 1     | 1     | 2     | 0     | 1     | 0.1571  | 6986464  | 9     | missense  | 0.152    | -0.084       | KDM4C | 5      |
| rs913588  | 0     | 0     | 0     | 1     | 0     | 0     | 0     | 1     | 0.3945  | 7174673  | 9     | missense  | 0.278    | 4.74         | KDM4C | 2      |
| rs1127430 | 0     | 0     | 1     | 0     | 0     | 0     | 1     | 0     | 0.4084  | 7799653  | 9     | missense  | 0.879    | 3.45         | C9orf123| 2    |
| rs35929428 | 0     | 0     | 0     | 1     | 0     | 0     | 0     | 1     | 0.08704 | 485834   | 9     | missense  | 0.321    | 4.49         | PTPRD | 2      |
| rs10961700| 0     | 1     | 0     | 1     | 1     | 1     | 0     | 1     | 0.2283  | 14776140 | 9     | missense  | 0.44     | 2.73         | FREM1 | 5      |
| rs35870000| 0     | 0     | 0     | 0     | 1     | 1     | 0     | 0     | 0.3655  | 14801710 | 9     | missense  | 0.003    | 1.79         | FREM1 | 2      |
| rs16932300| 0     | 1     | 0     | 1     | 0     | 0     | 0     | 1     | 0.06669 | 14801738 | 9     | missense  | 0.002    | -2.39        | FREM1 | 3      |
| rs7023244 | 0     | 1     | 1     | 1     | 0     | 0     | 0     | 1     | 0.28    | 14819370 | 9     | missense  | 0.995    | 5.86         | FREM1 | 4      |
| rs2779500 | 2     | 1     | 1     | 1     | 1     | 0     | 1     | 1     | 0.4838  | 14846036 | 9     | missense  | 0.012    | 0.988        | FREM1 | 7      |
| rs1407977 | 0     | 0     | 1     | 0     | 0     | 0     | 1     | 0     | 0.3497  | 15188106 | 9     | missense  | 0.002    | 2.85         | TTC39B| 2      |

‘0’ refers to a homozygous carrier of the major allele, ‘1’ refers to a heterozygous carrier and ‘2’ refers to a homozygous carrier of the minor allele.

MAF: Minor allele frequency
Supplementary Figure Legends

Supplementary Figure 1: Results of genome-wide linkage analysis. Green, red lines depicting the results of linkage analysis under dominant and recessive models respectively while the blue line shows the results of non-parametric linkage analysis.

Supplementary Figure 2: Result of haplotype analysis for chromosome 9p22-24. Affected individuals are marked in solid black. Solid grey individuals are un-typed individuals of unknown phenotype. Shared segregating haplotype is marked in red. All affected individuals are carriers of the 11 megabase long haplotype. Two brothers in the second-last generation are homozygous carriers of the haplotype.

Supplementary Figure 3: Result of haplotype analysis for chromosome 15q21. Two segregating haplotypes are marked in red and blue. Of 8 affected individuals in the two contributing families 5 are homozygous carriers of either the red or the blue haplotype, while two are compound heterozygous carriers of the red and blue haplotypes.

Supplementary Figure 4: Result of haplotype analysis for chromosome 19q13. 15 affected individuals in two contributing families. 8 are homozygous carriers of the haplotype and 7 are heterozygous carriers.

Supplementary Figure 5: Result of haplotype analysis for chromosome 20p13. Two contributing families. 10 affected individuals show homozygous sharing of the red haplotype and three appear to be compound heterozygotes.

Supplementary Figure 6: Result of haplotype analysis for chromosome 21q22. Segregating haplotype marked in red.
Supplementary Figure 7: Result of fine mapping for chromosome 9p22-24. c.1114C>T variant shared by 6 of the 8 haplotype carriers in Supplementary Figure 2. Two homozygous haplotype carriers are homozygous (T/T) for the mutation and treated for major depression in addition to high scores on HADS. Two heterozygous carriers (C/T) of the mutation are treated for mild depression while one is treated for stress and burnout.

Supplementary Figure 8: Family for the 9p22-24 region extended and results of Sanger sequencing. Solid black individuals were included in the initial linkage analysis. Other relatives are added by adding two more generations to the top. Three additional carriers of the mutation identified in blue, two of which have a high score on HADS (one is treated for mild depression). Of interest is an additional sibling (individual 38755) identified of the two homozygous carrier brothers. This sister unlike the other three siblings is not a carrier of the mutation (C/C) and has a low score on depression.

Supplementary Figure 9: RCL1 localization (A) RCL1 labeling was observed in the majority of nuclei as shown by co-localization with DAPI (scale bar = 20 µm). (B) RCL1 is densely present in neuronal nuclei of MAP2-positive neurons, as well as in MAP2-negative glial nuclei.

Supplementary Figure 10: RCL1-positive extensions are not neuronal. Labeling of RCL1 and neuron-specific markers reveals a lack of co-localization of RCL1 with (A) neuronal dendrites shown by MAP2, or (B) axons shown by SMI31 (scale bars = 10 µm).

Supplementary Figure 11: RCL1-positive extensions are CD44 positive. CD44-positive interlaminar extensions exhibit strong co-localization with RCL1 (scale bar = 10 µm).
Supplementary Figure 12: RCL1 marks a subset of GFAP-positive astrocytes. In contrast to interlaminar astrocyte processes at the cortical edge, GFAP-positive protoplasmic astrocytes in deeper cortical layers (layer V, VI) do not express cytoplasmic RCL1 (scale bars = 20 µm).

Supplementary Figure 13: RCL1 labeling of astrocytes is not observed in mouse brain. Adult mouse brain coronal sections of three different mice were stained for RCL1 and GFAP, which demonstrates the absence of cytoplasmic expression of RCL1 in astrocytes or any other cell type (dashed line marks the approximate border between cortical layers I and II, scale bars = 25 µm).

Supplementary Figure 14: Single immunohistochemical labeling of RCL1. In parallel with the double labeling experiments shown previously, single immunohistochemical labeling of RCL1 was performed to control for the specificity of the antibody labeling. RCL1 labeling was clearly present in nuclei (A) and interlaminar astrocyte extensions (B), including consideration of inherent tissue autofluorescence (AF) (scale bars = 20 µm).

Supplementary Figure 15: Negative controls of human postmortem immunohistochemical labeling. Sections were labeled with distinct secondary antibodies (A-C) or with all secondary antibodies together while omitting any primary antibodies (D) to control for the possibility of false positive or nonspecific labeling (scale bars = 20 µm).

Supplementary Figure 16: Ethnicity information provided by the principal components analysis. Black circles represent the samples analyzed, red: Caucasian, blue: Chinese, green: African, yellow: Colombian (AMR), cyan: Mexican (AMR), coral: Puerto Rican (AMR), brown: African (AMR).
Supplementary Figure 1

Depression

Chromosomes

LOD/HLOD

S

Supplementary Figure 1
