**Supplementary table 1.**

**Neurocognitive tests**

|  |  |  |
| --- | --- | --- |
| **Test** | **Additional testing information** | **Data interpretation** |
| *Intellectual functioning* The Dutch version of the Wechsler Adult Intelligence Scale-IV (WAIS-IV) (21). | Eight out of ten subtests of the WAIS-IV to estimate the four indexes: Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI) and Processing Speed Index (PSI)  The Verbal Comprehension Index (VCI) was estimated using the subtests of Similarities and Vocabulary. The Perceptual Reasoning Index (PRI) was estimated using the subtests Block Design and Matrix Reasoning. The Working Memory Index (WMI) was computed using the subtests Digit span and Arithmetic. The Processing Speed Index (PSI) was computed using the Symbol Search and Coding subtests. | The raw scores of the four indices of the WAIS-IV indexes were converted to age-related norm scores (M=100, SD=15). Estimated FSIQ scores were classified using the following categories [1] <70  (extremely low), [2] 70-85 (borderline), [3] 85-115 (average), [4] 115-130 (high), and [5] >130 (extremely high). |
| *Reaction times* FePsy (23). | Reaction times were quantified via the computerized task (FePsy) yielding simple visual and auditory reaction times of the dominant hand. | Scores of both tasks are expressed by speed of reaction time of the dominant hand in milliseconds (mean and SD of speed of reaction time). |
| *Executive functions* The Groninger Intelligence test-2 (GIT-2) (25). | a subtest of the Groninger Intelligence test-2 (GIT-2) was used to assess word fluency  During this test, participants had to name as many animals (trial 1: word fluency animals) and professions (trial 2: word fluency professions) as possible within one minute. |  |
| *Executive functions:* The Zoo Map subtest of the Behavioral Assessment of Dysexecutive Syndrome (BADS) battery (24). | This test was used to evaluate planning skills.  The first high demand trial assesses spontaneous planning whereas the second low demand trial assesses the ability of participants to follow a preformulated set of instructions. | Outcomes of both trials yielded a total raw score of correctly visited locations in an accurate order with points subtracted for errors (maximum score per trial=16). |
| *Memory* Rey auditory verbal learning test (RAVLT) (26). | At testing, a list of 15 unrelated meaningful words were presented five times, each time in the same order and after each trial the number of correctly recalled words were measured (direct recall). Participants had to recall as many words as possible after an interval period of 20 minutes (delayed recall). Finally, a list of 30 words was presented and participants were asked to confirm whether they recognized the 15 unrelated previous words out of the 30 new words presented (recognition recall). | Scores were computed to (1) a sum of correct responses given during the five consecutive trials (total immediate recall score), (2) total correct responses after a delay (delayed recall score), (3) sum of correct recognition responses (recognition trial) and (4) learning score of correct responses on each immediate recall trial (learning curve of trials 1 to 5).  The RAVLT has an ‘embedded indicator’ to evaluate symptom validity. Malingering was checked by computing the total score on recognition and delayed recall. The score on recognition should be higher than delayed recall score +15. |

**Behavioral self-report questionnaires**

|  |  |
| --- | --- |
| **Questionnaire** | **Data interpretation** |
| Behavior Rating Inventory of Executive Function for adults (BRIEF-A) (27). | The BRIEF-A is a 75-item questionnaire capturing self-reported everyday executive functioning.  Responses were scored as never [1] to often [3]. Results yield two index scores, the Metacognition index (MCI) and Behavioral Regulation Index (BRI) and one overall Global Executive Composite score (GEC). The MCI index consist of the subscales: working memory, initiate, planning, task monitoring and organization of materials. The BRI consists of the subscales inhibition, shifting, emotional control and self-monitoring. Raw scores for each scale are summed and transformed to age-corrected T-scores (mean = 50 and SD = 10). |
| Strengths and Difficulties Questionnaire (SDQ) (28). | The SDQ screens for mental health (behavioral and emotional).  It consists of 28 items rated on a 3-point scale (0 = not true to 2 = certainly true) and assessing five domains: [1] emotional symptoms, [2] conduct problems, [3] hyperactivity/inattention, [4] peer relationship problems and [5] prosocial behavior. All subscales except the prosocial behavioral subscale were summed to calculate the total difficulties score, with higher scores indicating more mental health difficulties. |
| Hospital Anxiety and depression Scale (HADS) (29). | The HADS assess anxiety and depression in somatic and psychiatric patients.  All items were rated on a 4-point scale. A HADS score ranging from 0 to 7 indicates the absence of anxiety or depressive symptoms, a score between 8 to 11 indicates mild to moderate symptoms, and scores greater than 12 indicates a significant number of symptoms suggesting a clinical diagnosis of anxiety or depression. |
| The Dutch Symptom CheckList-90-Revised (SCL-90-R) (30). | The SCL-90-R consists of 90 items covering eight domains of psychological symptoms: anxiety, depression, agoraphobia, somatization, cognitive performance deficits, interpersonal sensitivity, hostility and sleep difficulties.  Items were rated on a 5-point scale (1 = no complaints at all to 5 = high level of complaints) and summed to a total score and eight subtest-scores, with higher scores reflecting more psychological distress. |

**Supplementary table 2.**

**Genetic mutations of study participants**

