ARTICLE

https://doi.org/10.1038/s41586-019-1195-2

Single-cell transcriptomic analysis of Alzheimer's disease

Hansruedi Mathys[1,2,10], Jose Davila-Velderrain[3,4,10], Zhuyu Peng[1,2], Fan Gao[1,2], Shahin Mohammadi[3,4], Jennie Z. Young[1,2], Madhvi Menon[4,5,6], Liang He[3,4], Fatema Abdurrob[1,2], Xueqiao Jiang[1,2], Anthony J. Martorell[1,2], Richard M. Ransohoff[7], Brian P. Hafler[4,5,6,8], David A. Bennett[9], Manolis Kellis[3,4,11]* & Li-Huei Tsai[1,2,4,11]*

[1]Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, USA. [4]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [5]Department of Neurology, Harvard Medical School, Boston, MA, USA. [6]Evergrande Center for Immunologic Diseases, Harvard Medical School, Boston, MA, USA. [7]Third Rock Ventures, Boston, MA, USA. [8]Department of Ophthalmology, Harvard Medical School, Boston, MA, USA. [9]Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, USA. [10]These authors contributed equally: Hansruedi Mathys, Jose Davila-Velderrain. [11]These authors jointly supervised this work: Manolis Kellis, Li-Huei Tsai. *e-mail: manoli@mit.edu; lhtsai@mit.edu

Description of variables according to the Rush Alzheimer's Disease Center Codebook

- **age_death - Age at death**

- **amyloid - Overall amyloid level - Mean of 8 brain regions**
  Amyloid beta protein identified by molecularly-specific immunohistochemistry and quantified by image analysis. Value is percent area of cortex occupied by amyloid beta. Mean of amyloid beta score in 8 regions (4 or more regions are needed to calculate). The 8 regions are hippocampus, entorhinal cortex, midfrontal cortex, inferior temporal gyrus, angular gyrus, calcarine cortex, anterior cingulate cortex, superior frontal cortex.

- **braaksc - Braak stage**
  Semi quantitative measure of neurofibrillary tangles.
  Braak Stage is a semi quantitative measure of severity of neurofibrillary tangle (NFT) pathology. Bielschowsky silver stain was used to visualize NFTs in the frontal, temporal, parietal, entorhinal cortex, and the hippocampus. Braak stages were based upon the distribution and severity of NFT pathology: Braak stages I and II indicate NFTs confined mainly to the entorhinal region of the brain; Braak stages III and IV indicate involvement of limbic regions such as the hippocampus; Braak stages V and VI indicate moderate to severe neocortical involvement.

- **ceradsc - CERAD score**
  Semiquantitative measure of neuritic plaques.
  CERAD score is a semiquantitative measure of neuritic plaques. A neuropathologic diagnosis was made of no AD (value 4), possible AD (value 3), probable AD (value 2), or definite AD (value 1) based on semiquantitative estimates of neuritic plaque density as recommended by the Consortium to Establish a Registry for Alzheimer's Disease (CERAD), modified to be implemented without adjustment for age and clinical diagnosis. A CERAD neuropathologic diagnosis of AD required moderate (probable AD) or frequent neuritic plaques (definite AD) in one or more neocortical regions. Diagnosis includes algorithm and neuropathologist's opinion, blinded to age and all clinical data.

  Value 1: definite AD
  Value 2: probable AD
  Value 3: possible AD
  Value 4: no AD

- **cogdx - Final consensus cognitive diagnosis**
  Clinical consensus diagnosis of cognitive status at time of death.
  At the time of death, all available clinical data were reviewed by a neurologist with expertise in dementia, and a summary diagnostic opinion was rendered regarding the most likely clinical diagnosis at the time of death. Summary diagnoses were made blinded to all postmortem data. Case conferences including one or more neurologists and a neuropsychologist were used for consensus on selected cases.

Value 1: NCI, No cognitive impairment (No impaired domains)
Value 2: MCI, Mild cognitive impairment (One impaired domain) and NO other cause of CI
Value 3: MCI, Mild cognitive impairment (One impaired domain) AND another cause of CI
Value 4: AD, Alzheimer's disease and NO other cause of CI (NINCDS PROB AD)
Value 5: AD, Alzheimer's disease AND another cause of CI (NINCDS POSS AD)
Value 6: Other dementia. Other primary cause of dementia

- **cogn_global - Global cognitive function**
  Global cognitive function - Average of 19 tests.
  cogn_global_lv ("lv" means last valid score)
  Cogn_global is the main variable for overall (i.e. global) cognitive function. Raw scores from a battery of cognitive tests were converted to Z scores and averaged to yield a global cognitive function summary.

- **dlbdx - Lewy Body disease**
  Pathologic diagnosis of Lewy body diseases - 4 stages.
  Pathologic diagnosis of Lewy Body disease describes 4 stages of distribution of α-synuclein in the brain based on algorithm and neuropathologist's opinion.

  Value 0: not present
  Value 1: nigral-predominant
  Value 2: limbic-type
  Value 3: neocortical-type

- **educ – Education**
  Years of education. The years of education variable is based on the number of years of regular school reported at baseline cognitive testing.

- **gpath - Global AD pathology burden**
  Global burden of AD pathology based on 5 regions.
  Global AD pathology burden is a quantitative summary of AD pathology derived from counts of three AD pathologies: neuritic plaques (n), diffuse plaques (d), and neurofibrillary tangles (nft), as determined by microscopic examination of silver-stained slides from 5 regions: midfrontal cortex (midfrontal), midtemporal cortex (midtemp), inferior parietal cortex (infparietal), entorhinal cortex (ento), and hippocampus (ca1hip). Each regional count is scaled by dividing by the corresponding standard deviation. The 5 scaled regional measures for each type of pathology are then averaged to obtain summary measures (plaq_d, plaq_n, and nft). The 3 summary measures are then averaged to obtain the measure of global AD pathology.

- **gpath_3neocort - Global measure of neocortical pathology based on scaled scores**
  This is calculated by taking the mean of 3 scaled variables from regions (midfrontal, midtemp, infparietal), where the scaled variable is the original count divided by the standard deviation.

- **msex – Sex**
  Self-reported sex, with “1” indicating male sex.
  1 = Male
  0 = Female

- **nft - Neurofibrillary tangle burden**
  Neurofibrillary tangle summary based on 5 regions.
  Neurofibrillary tangle burden is determined by microscopic examination of silver-stained slides from 5 regions: midfrontal cortex, midtemporal cortex, inferior parietal cortex, entorhinal cortex, and hippocampus. The count of each region is scaled by dividing by the corresponding standard deviation. The 5 scaled regional measures are then averaged to obtain a summary measure for neurofibrillary tangle burden.

- **niareagansc - NIA-Reagan score**
  NIA Reagan Diagnosis of Alzheimer's disease - 4 levels (none to high likelihood).
  The modified NIA-Reagan score based on consensus recommendations for postmortem diagnosis of Alzheimer’s disease. The criteria rely on both neurofibrillary tangles (Braak) and neuritic plaques (CERAD).

  Value 1: High likelihood of AD
  Value 2: Intermediate likelihood of AD
  Value 3: Low likelihood of AD
  Value 4: no AD

- **parksc - Parkinsonian signs**
  Global parkinsonian summary score.
  The global parkinsonian summary score is a composite measure of parkinsonian signs. The score is the average of 4 separate domains calculated from a 26-item modified version of the motor portion of the United Parkinson’s Disease Rating Scale (mUPDRS): Bradykinesia, Gait, Rigidity, and Tremor. The global parkinsonian summary score ranges from 0 to 100 and is calculated by averaging the four domain scores. A higher summary score reflects expression of more severe Parkinsonian signs.

- **plaq_d - Diffuse plaque burden**
  Diffuse plaque summary based on 5 regions.
  Diffuse plaque burden is determined by microscopic examination of silver-stained slides from 5 regions: midfrontal cortex, midtemporal cortex, inferior parietal cortex, entorhinal cortex, and hippocampus. The count of each region is scaled by dividing by the corresponding standard deviation. The 5 scaled regional measures are then averaged to obtain a summary measure for diffuse plaque burden.

- **plaq_n - Neuritic plaque burden**
  Neuritic plaque summary based on 5 regions.
  Neuritic plaque burden is determined by microscopic examination of silver-stained slides from 5 regions: midfrontal cortex, midtemporal cortex, inferior parietal cortex, entorhinal cortex, and hippocampus. The count of each region is scaled by dividing by the

corresponding standard deviation. The 5 scaled regional measures are then averaged to obtain a summary measure for neuritic plaque burden.

- **pmi - Post-mortem interval**
  Time interval in hours from time of death to autopsy.
  Post-mortem interval (PMI) refers to the interval between death and tissue preservation in hours.

- **tangles – Tangles**
  Tangle density - Mean of 8 brain regions.
  Neuronal neurofibrillary tangles are identified by molecularly specific immunohistochemistry (antibodies to abnormally phosphorylated Tau protein, AT8). Cortical density (per mm2) is determined using systematic sampling. Mean of tangle score in 8 regions (4 or more regions are needed to calculate). The 8 regions used are hippocampus, entorhinal cortex, midfrontal cortex, inferior temporal gyrus, angular gyrus, calcarine cortex, anterior cingulate cortex, superior frontal cortex.