A bird's eye view to the homeostatic, Alzheimer and Glioblastoma attractors

Joan Nieves, Gabriel Gil and Augusto Gonzalez

Institute of Cybernetics, Mathematics and Physics, Havana

A well known paradigm in molecular genetics expresses that local maxima of fitness in gene expression space are related to biological viable states [1]. This picture has been applied to the description of cell fates along differentiation lines [2]. However, to the best of our knowledge, there are no plots based on real data for a given tissue representing at least a partial landscape with more than two of these maxima. We provide a drawing for white matter of the brain in which the normal state is represented along with the glioblastoma (GBM) attractor and the seemingly modest maximum related to Alzheimer disease (AD). The plot shows that GBM and AD are opposite alternatives, as epidemiological [3,4,5,6] and molecular biology studies [7,8,9] suggest, and indicates a path or corridor for normal aging, in accordance with the programmatic aging theory [10,11]. At the gene level, there are genes varying in the same way in the aging, AD progression and cancer processes, whereas there are also genes indicating the disjunctive between AD and GBM. An example of the latter is the MMP9 protein-coding gene, playing an important role in tumor invasion [12,13], but known also as a neuroprotector, controlling the interactions between axons and beta-amyloid fibers [14]. Deviations of the gene expression value from its reference in normal tissue may indicate either a potential progression to AD (under-expression) or to GBM (over-expression). This unusual view, following from a simple plot, may help understand the relations between AD and GBM biology and identify useful gene markers for both processes. We anticipate that similar diagrams could be important in other tissues.

The N + GBM + AD diagram.

Our starting point is the principal component analysis **[15]** diagram of gene expression data for white matter of the brain, shown in **Fig. 1a)**. Four groups of samples are apparent in this figure. Samples labeled as N and GBM correspond, respectively, to pathologically normal and tumor specimens in The Cancer Genome Atlas data for Glioblastoma (TCGA, https://www.cancer.gov/tcga) **[16]**. They are taken during surgery procedures. Tumors are geographically localized in different brain zones but, as it is common for Glioblastoma, they are white matter tumors **[17]**. The centers of the N and GBM clouds of samples in gene expression space define, respectively, the Normal (homeostatic) and Glioblastoma Kaufmann attractors **[18,19]**.

On the other hand, the groups labeled as AD and O correspond, respectively, to Alzheimer disease and control white matter samples in the Allen Institute study on aging and dementia (http://aging.brain-map.org/) **[20]**. They are taken post mortem. The O group comes from normal aged patients, with ages ranging in the interval between 77 and 101 years. We studied the O to AD transition in Ref. **[21]**. As age increases, we observe a displacement of O samples towards the center of the AD cloud. Average positions of AD subgroups of samples, however, are fixed irrespective of age. This property is apparent in Fig. 4 of Ref. [21]. From these facts, we conclude that the center of the AD cloud of samples may define an attractor in gene expression space, whereas O samples are captured by the AD attractor in the process of aging.

Technical difficulties arising when trying to use two such unalike clinical studies as sources of data for a common principal component analysis calculation are partially detailed in the Supplementary Information. The ideal situation would be to repeat the studies within a unique technological framework, and to include data from young normal people, which should be used to set the reference

for differential gene expression calculations, to include data from GBM and AD patients, and data for normal patients in different age ranges. We look at our **Fig 1 a)** diagram as a qualitative approximation to this ideal experiment.

Thus, in our approximation we get a gene expression space landscape with 3 attractors: N, GBM and AD, and a set of O samples moving towards the latter. The relative positions and main transitions between attractors are summarized in **Fig. 1b)**. We assume that they are determined by the Biology underlying the processes in the tissue. The N to AD transition is labeled as "early onset of AD" in order to stress that there is also a way to AD through aging, the "late onset of AD". A path for aging is also signaled in the figure. We shall come back to this point below.

There is still an additional qualitative information which can be introduced in our scheme. It is related to a fitness variable, in such a way that we draw a kind of Wright diagram **[1]**. A cartoon containing a contour plot of fitness is represented in **Fig. 1c)**. The N and GBM attractors are fitness maxima, and they should be separated by a low-fitness barrier **[21]**. The GBM should be the highest maximum **[21,22]**. On the other hand, the transition from O to AD is quasi-continuous, with a relatively small number of differentially expressed genes **[21]**. It means that there is a very small barrier or even a barrier-free path connecting O and AD. We expect a low-fitness barrier preventing the direct transitions from N to AD, and a small AD maximum, as this attractor is located in the far from N low-fitness region. All of these facts are represented in **Fig. 1c)**. The cartoon is constructed from a sum of Gaussians centered at the attractors, with standard deviations proportional to the actual values observed in **Fig. 1a)**, and with heights respecting the relative strengths of attractors.

Main results.

On the basis of our diagrams, we may formulate the following remarks or statements, which are the main results of the paper:

Statement 1. There is a direction in gene expression space, which roughly speaking may be identified with the PC1 axis, associated with aging and with an increase in the risk for AD and GBM.

Indeed, displacement along this direction implies partially climbing the low-fitness barriers separating N from the AD and GBM states, and thus augmenting the risk for both AD and GBM.

It is worth looking at the main genes involved in this process. To this end, we look at the unitary vector along the PC1 axis. Genes are ranked according to their contribution to the vector. The procedure is similar to the Page Rank algorithm **[23]**. We used it in our previous work **[19]**. A list with the first 100 genes in the ranking is given in **Supplementary Table I**. Positive amplitudes defines genes which expression increases in the displacement along the positive direction of PC1, whereas negative amplitudes refer to silenced genes**.** These genes should simultaneously play a crucial role in aging, GBM and AD.

Of course, due to the qualitative-only value of our analysis, the genes and specially the ranking should be taken with care. Nevertheless, notice the presence of RNA genes, and particularly micro-RNA genes, in the Table, the role of which is widely recognized in recent literature **[24,25]**.

Statement 2. There is a direction in gene expression space, which may be roughly identified with the PC2 axis, showing that AD and GBM are excluding alternatives.

Indeed, AD and GBM seem to be in opposite semiplanes. Clinical evidence **[3,4,5,6]** and molecular biology studies **[7,8,9]** support this disjunctive. Consequently, the PC2 axis involve genes inversely deregulated in AD and GBM. In **Supplementary Table II**, we list the top 100 genes defined by the unitary vector along the PC2 axis. Positive weights correspond to genes which expression increases in the N to AD transition. On the other hand, negative amplitudes correspond to genes with increasing expression in the N to GBM transition. The latter are related to regulation of inflammation, ECM remodeling, regulation of translation, Glucose metabolism, inhibition of apoptosis, cell cycle and the regulation of transcription, i.e. to cancer hallmarks **[26]**.

Above, we mentioned MMP9 as an example of these genes. The UBE2C protein-coding gene is another example **[27,28]**. **Fig. 1d)** shows violin plots for the differential expression of both genes in N, AD and GBM samples. They are over-expressed in the N to GBM transition, but silenced in the early N to AD transition. Notice also in **Supplementary Table II** the presence of long non-coding and micro-RNA genes, inversely regulated in both processes.

<u>Statement 3</u>. There is an aging corridor, that is a preferential path for aging in gene expression space.

This statement comes from the location of the O set of points and the qualitative distribution of fitness in **Fig. 1c)**. Naively, one would expect for aging a kind of Brownian motion starting in the N region **[29]**, which should lead to a uniform nearly spherical distribution for aged normal samples. Instead of that, samples corresponding to normal aged brains are located in a definite region of gene expression space, close to the AD attractor. In other words, the process of aging seems to define a trajectory or corridor of continuously decreasing fitness, from which the O data shows the last segment.

This fact is consistent with the hypothesis of programmatic aging **[10,11]**, i.e. the idea that aging is programmed in our genes. The "program" could be the result of Darwinian selection in the fitness landscape sketched in **Fig. 1d)**. It is widely accepted the role of Darwinian selection in somatic evolution **[30]**. The corridor is preferred against directions in gene expression space with still lower fitness. Many positive selection events could lead to fixation in our genes.

Notice also that the corridor is a way of avoiding the strong GBM attractor.

Statement 4. The late onset of AD is the result of capture by the AD attractor of aged brain micro states.

The picture is, thus, as follows. The aging corridor defines a smooth path connecting N and AD. The process of aging is initially related to a displacement along the positive direction of the PC1 axis with the corresponding decrease of fitness. In the last steps, the O states are captured by the weak AD attractor, implying a relative increase of fitness, possibly improving energy intake and metabolism at the cost of decreasing brain functionality.

This statement is supported by the list in **Supplementary Table III** of the top 10 genes in the O to AD transition. They come from calculations in Ref. **[21]**. We included a column with the phenotypes associated to these genes, according to GWAS **[31]**. Most of the genes are indeed related to metabolism, energy intake and, in general, the increase of fitness.

Final remarks.

Our simple qualitative drawings identify directions in gene expression space associated to different biological processes: AD onset, carcinogenesis, aging. Every one of these directions is characterized by

a “metagene” or gene expression profile, from which the main genes contributing to the process can be extracted.

Some of our results confirm previous knowledge, but others require further corroborations. For example, the idea that programmatic aging could be related to avoiding the strongest GBM attractor, or the late onset of AD as the capture by the AD attractor of normal aged samples. We hope, they will motivate experimental research work along these directions.

Let us stress that even more refined data or computational methods could not essentially modify our qualitative schemes with only 3 attractors. Their relative positions could vary, but the formulated statements will remain.

Fig. 1a) should be completed with data corresponding to other kinds of dementia or brain disorders. In particular, one should expect a Parkinson disease area close to the AD attractor and opposite to GBM **[32]**. The whole picture may reveal a still finer topology of gene expression space and a richer Wright diagram.

We anticipate that similar diagrams in other tissues, besides providing an integral perspective, could be useful in the understanding of the biology of apparently unrelated diseases or disorders, and in the discovery of unexpected clues for their treatment.

References

[1] Sewall Wright (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. Proceedings of the 6th International Congress on Genetics 1: 356–366.

[2] Michael J. Casey, Patrick S. Stumpf, and Ben D. MacArthur (2020). Theory of cell fate. Wiley Interdiscip. Rev. Syst. Biol. Med. 12(2): e1471. https://doi.org/10.1002/wsbm.1471

[3] Shuo-Ming Ou, Yi-Jung Lee, Yu-Wen Hu et al (2013). Does Alzheimer's Disease Protect against Cancers? A Nationwide Population-Based Study. Neuroepidemiology 40: 42–49. https://doi.org/10.1159/000341411

[4] Jane A Driver, Alexa Beiser, Rhoda Au, et al (2012). Inverse association between cancer and Alzheimer's disease: results from the Framingham Heart Study. BMJ 344: e1442. https://doi.org/10.1136/bmj.e1442

[5] C.M. Roe, A.L. Fitzpatrick, C. Xiong, et al (2010). Cancer linked to Alzheimer disease but not vascular dementia. Neurology 74: 106 – 112. https://doi.org/10.1212/WNL.0b013e3181c91873

[6] Massimo Musicco, Fulvio Adorni, Simona Di Santo et al (2013). Inverse occurrence of cancer and Alzheimer disease: a population-based incidence study. Neurology 81(4): 322-328. https://doi.org/10.1212/WNL.0b013e31829c5ec1

[7] Timothy Liu, Ding Ren, Xiaoping Zhu et al (2013). Transcriptional signaling pathways inversely regulated in Alzheimer's disease and glioblastoma multiform. Sci. Rep. 3: 3467. https://doi.org/10.1038/srep03467

[8] Cristina Lanni, Mirco Masi, Marco Racchi and Stefano Govoni (2020). Cancer and Alzheimer's disease inverse relationship: an age-associated diverging derailment of shared pathways. Molecular Psychiatry 26: 280–295. https://doi.org/10.1038/s41380-020-0760-2

[9] Jiayang Cai, Liguo Ye, Yuanyuan Hu et al (2022). Exploring the inverse association of glioblastoma multiforme and Alzheimer's disease via bioinformatics analysis. Medical Oncology 39: 182. https://doi.org/10.1007/s12032-022-01786-w

[10] João Pedro de Magalhães (2012). Programmatic features of aging originating in development: aging mechanisms beyond molecular damage? FASEB J. 26(12): 4821–4826.

https://doi.org/10.1096/fj.12-210872

[11] David Gems (2022). The hyperfunction theory: an emerging paradigm for the biology of aging. Ageing Res Rev. 74: 101557. https://doi.org/10.1016/j.arr.2021.101557

[12] Gheeyoung Choe, Jun K Park, Lisa Jouben-Steele, et al (2002). Active matrix metalloproteinase 9 expression is associated with primary glioblastoma subtype. Clin Cancer Res 8(9): 2894-2901.

[13] Qiang Xue, Li Cao, Xiao-Yan Chen, et al (2017). High expression of MMP9 in glioma affects cell proliferation and is associated with patient survival rates. Oncol Lett 13(3): 1325–1330. https://doi.org/10.3892/ol.2017.5567

[14] Archontia Kaminari, Nikolas Giannakas, Athina Tzinia & Effie C. Tsilibary (2017). Overexpression of matrix metalloproteinase-9 (MMP-9) rescues insulin-mediated impairment in the 5XFAD model of Alzheimer's disease. Sci. Rep. 7: 683. https://doi.org/10.1038/s41598-017-00794-5

[15] Lever, J., Krzywinski, M., Altman, N. (2017). Principal component analysis. Nature Methods 14(7): 641–642. https://doi.org/10.1038/nmeth.4346

[16] Cameron W. Brennan, Roel G.W. Verhaak, Aaron McKenna, et al (2013). The Somatic Genomic Landscape of Glioblastoma. Cell 155: 462–477. http://dx.doi.org/10.1016/j.cell.2013.09.034

[17] Ellingson B.M., Lai A., Harris R.J., et al (2013). Probabilistic radiographic atlas of glioblastoma phenotypes. AJNR Am J Neuroradiol. 34(3): 533-540. https://doi.org/10.3174/ajnr.A3253

[18] Sui Huang, Ingemar Ernberg, and Stuart Kauffman (2009). Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective. In: Seminars in Cell & Developmental Biology 20: 869–876. doi: https://doi.org/10.1016/j.semcdb.2009.07.003

[19] Gonzalez, A., Leon, D.A., Perera, Y., Perez, R.: On the gene expression landscape of cancer (2020). PLoS ONE 18(2): e0277786. https://doi.org/10.1371/journal.pone.0277786

[20] Jeremy A. Miller, Angela Guillozet-Bongaarts, Laura E. Gibbons, et al (2017). Neuropathological and transcriptomic characteristics of the aged brain. ELife 6: 31126. https://doi.org/10.7554/eLife.31126

[21] Augusto Gonzalez, Joan Nieves, Dario A. Leon-Valido et. al. (2021). Gene expression rearrangements denoting changes in the biological state. Sci. Rep. 11: 8470. https://doi.org/10.1038/s41598-021-87764-0

[22] Augusto Gonzalez, Frank Quintela, Dario A. Leon, et al (2022). Estimating the number of available states for normal and tumor tissues in gene expression space. Biophysical Reports 2: 100053. https://doi.org/10.1016/j.bpr.2022.100053

[23] N. Duhan, A. K. Sharma and K. K. Bhatia (2009). Page Ranking Algorithms: A Survey. IEEE International Advance Computing Conference, Patiala, India. pp. 1530-1537. https://doi.org/10.1109/IADCC.2009.4809246

[24] Saverio Candido, Gabriella Lupo, Manuela Pennisirna et al (2019). The analysis of miRNA expression profiling datasets reveals inverse microRNA patterns in Glioblastoma and Alzheimer's disease. ONCOLOGY REPORTS 42: 911-922. https://doi.org/10.3892/or.2019.7215

[25] Laura Thomas, Tullio Florio and Carolina Perez-Castro (2020). Extracellular vesicles loaded miRNAs as potential modulators shared between Glioblastoma, and Parkinson’s and Alzheimer’s Diseases. Front. Cell. Neurosci. 14: 590034. https://doi.org/10.3389/fncel.2020.590034

[26] Hanahan D., Weinberg R.A. (2000). The Hallmarks of Cancer. Cell. 100(1): 57–70. https://doi.org/10.1016/S0092-8674(00)81683-9

[27] RUIMIN MAI, XIXIONG KANG, GUOJUN ZHANG, et al (2016). High expression of UBE2C is associated with the aggressive progression and poor outcome of malignant glioma. ONCOLOGY LETTERS 11: 2300-2304. https://doi.org/10.3892/ol.2016.4171

[28] Suraj K. Jaladanki, Abdulkadir Elmas, Gabriel Santos Malave & Kuan-lin Huang (2021). Genetic dependency of Alzheimer’s disease-associated genes across cells and tissue types. Scientific Reports 11: 12107. https://doi.org/10.1038/s41598-021-91713-2

[29] Roberto Herrero, Dario A. Leon & Augusto Gonzalez (2022). A one-dimensional parameter-free model for carcinogenesis in gene expression space. Scientific Reports 12: 4748. https://doi.org/10.1038/s41598-022-08502-8

[30] Di Gregorio A., Bowling S., Rodriguez T.A. (2016). Cell Competition and Its Role in the Regulation of Cell Fitness from Development to Cancer. Dev. Cell. 38(6): 621-634. https://doi.org/10.1016/j.devcel.2016.08.012

[31] Sollis E., Mosaku A., Abid A., et al. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Research, 51(D1): D977–D985. https://doi.org/10.1093/nar/gkac1010

[32] Pauline Mencke, Zoé Hanss, Ibrahim Boussaad, et al (2020). Bidirectional Relation Between Parkinson's Disease and Glioblastoma Multiforme. Front. Neurol. 11: 898. https://doi.org/10.3389/fneur.2020.00898

Acknowledgments. Authors acknowledge the Office of External Activities of the Abdus Salam Centre for Theoretical Physics for support. The research is carried on under a project of the Nuclear Energy and Advanced Technologies Agency (AENTA), Cuba.

Author's contributions.

A.G. conceived and coordinated the work. G.G. suggested the connection with programmatic aging. J.N. processed the data and is in charge of the GitHub repository. All authors analyzed and interpreted the results, contributed to the manuscript and approved the final version.

Competing interests.

The authors declare that they have no competing interests.

Availability of data and materials.

The data we used, the Python routines and the main results are integrated in the following public GitHub repository: https://github.com/JoanANievesCuadrado/GBM-ALZ

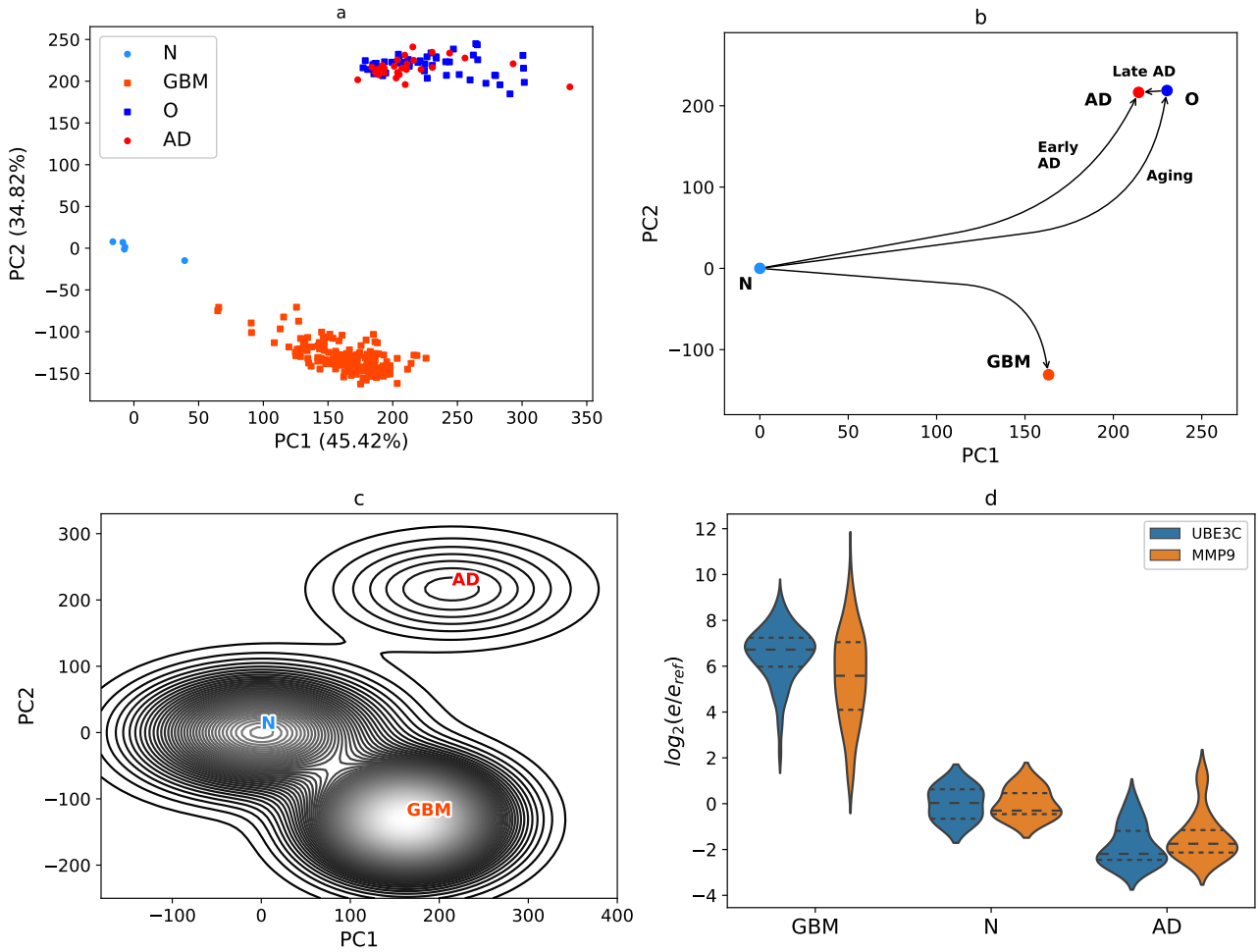

Fig. 1. a) Principal component analysis of the data studied in the paper. N – Normal homeostatic state, GBM – Glioblastoma, AD – Alzheimer disease state. The normal old samples are denoted by O. **b)** Schematics of the transitions between attractors. **c)** A Wright diagram showing a hypothetical contour plot of fitness. The absolute maximum corresponds to the GBM state. The AD attractor is represented as a slight local maximum. **d)** Violin plot for the log-fold changes of MMP9 and UBE2C genes in N, AD and GBM states. The geometric mean of the expression in the N state is taken as reference in order to compute differential expression values.

Supplementary Information

As mentioned, we use gene expression data, in FPKM format, from Refs. **[15,16]**. The data was obtained by using different platforms. We took the approximately 30,000 genes that are perfectly identified in both platforms and perform a simple Principal Component Analysis (PCA) **[17]**, described elsewhere **[18,19]**. The common reference used to define log-fold differential expression values and compute the covariance matrix for the PCA is the geometric mean in the N state. The 5 samples in the N state come from the TCGA data. There are also 169 GBM samples. On the other hand, in the Allen Institute data for white matter of the brain there are 47 control samples, which conform our old (O) group, and 28 AD samples.

There are both conceptual and technical issues arising when using these two dissimilar experiments in a single PCA calculation. For example, the reference N is not precisely the normal state, but a set of pathologically normal samples taken from individuals with GBM tumors. Two of the patients are even older than 70 years. From the computational side, on the other hand, one could use batch corrections **[33,34]**, which partially amend the biases associated to each group of samples, but may introduce also uncontrolled artifacts.

Thus, we decided to take the data as it is, and use the simplest PCA technique, without any sophistication. We don't believe that any correction will essentially change the qualitative analysis following from the 3-attractors diagrams shown in **Fig. 1**.

[33] Haghverdi, L., Lun, A., Morgan, M. et al (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol* **36:** 421–427. https://doi.org/10.1038/nbt.4091

[34] Zhang Y., Parmigiani G., Johnson W.E. (2020). *ComBat-seq*: batch effect adjustment for RNA-seq count data. NAR Genom Bioinform 2(3): lqaa078. **https://doi.org/10.1093/nargab/lqaa078**

Supplementary Tables.

Gene	Weight in PC1
RN7SL1	0.04
RPPH1	0.03
RNU2-1	0.03
SNORD3A	0.03
RNU1-2	0.03
RMRP	0.03
RNU1-4	0.03
RNU1-3	0.03
RNVU1-18	0.03
RNU1-1	0.03
SNORA73A	0.02
RNVU1-7	0.02
RNU2-2P	0.02
WARS2-IT1	0.02
…	…
VGF	-0.02
SLC12A5	-0.02
HIPK4	-0.02
SSTR1	-0.02
CPLX1	-0.02
SYT5	-0.02
MIR657	-0.02
INA	-0.02
CCK	-0.02
GABRD	-0.02
MIR1249	-0.02
WIF1	-0.02
NECAB2	-0.02
FBXL16	-0.02
GDA	-0.02
KIAA1045	-0.02
SYNPR	-0.02
HMP19	-0.02
CHGA	-0.02
SNCG	-0.02
SVOP	-0.02

LRTM2	-0.02
WNT10B	-0.02
MTND4P12	-0.02
PHYHIP	-0.02
RAB3A	-0.02
RASAL1	-0.02
HPCA	-0.03
SERTM1	-0.03
GABRG2	-0.03
CAMKV	-0.03
CREG2	-0.03
SNAP25	-0.03
GAD2	-0.03
SYT13	-0.03
KLK7	-0.03
WBSCR17	-0.03
SV2B	-0.03
SLC30A3	-0.03
ATP1A3	-0.03
CPNE6	-0.03
CAMK2A	-0.03
SLC6A7	-0.03
GABRA1	-0.03
KCNJ4	-0.03
CDK5R2	-0.03
CALY	-0.03
GNG3	-0.03
CALB2	-0.03
CHD5	-0.03
VSNL1	-0.03
CRYM	-0.03
CCKBR	-0.03
NEUROD6	-0.03
HRH3	-0.03
SYT1	-0.03
SNCB	-0.03
CHRM1	-0.03
MAL2	-0.03
SNORD113-3	-0.03
PNMA5	-0.03

CPLX2	-0.03
TMEM130	-0.03
LOC105373377	-0.03
SYN2	-0.03
SLC6A17	-0.03
ICAM5	-0.03
NEFL	-0.03
SYN1	-0.03
C1QL3	-0.03
NEFM	-0.03
DDN	-0.03
FAM163B	-0.03
GABRA5	-0.03
NRGN	-0.03
SLC32A1	-0.03
SLC8A2	-0.03
PACSIN1	-0.03
SULT4A1	-0.03
CACNG3	-0.03
GRIN1	-0.03
SLC17A7	-0.03
PDYN	-0.03
PRKCG	-0.03
SST	-0.04
MIR770	-0.04

Supplementary Table I. The top 100 relevant genes contributing to the unitary vector along PC1 in the common plot.

Gene	Weight in PC2
RN7SL1	0.05
RPPH1	0.04
RNU2-1	0.04
SNORD3A	0.04
RMRP	0.04
RNU1-3	0.03
RNU1-1	0.03
RNU1-4	0.03
RNVU1-18	0.03
RNU1-2	0.03
SNORA73A	0.03
RNU2-2P	0.03
BCYRN1	0.03
LOC101928075	0.03
RNY3	0.03
SNORA63	0.03
RNVU1-7	0.03
SCARNA2	0.03
SNORA48	0.02
CYCSP30	0.02
SCARNA10	0.02
RN7SL2	0.02
LOC729348	0.02
SCARNA5	0.02
SCARNA7	0.02
SNORA81	0.02
SNORD97	0.02
DNAJC19P8	0.02
FCF1P1	0.02
SNORA57	0.02
SNORD10	0.02
SNORA49	0.02
SNORA54	0.02
RAB9AP2	0.02
RAB9AP5	0.02
RNY1	0.02
…	…
TPT1P9	-0.02
RPL31P4	-0.02
RPL10P9	-0.02
GAPDHP65	-0.02
RPL12P4	-0.02
RPS20P14	-0.02
FAM64A	-0.02
RPS2P46	-0.02

RPS3AP5	-0.02
RPL7P1	-0.02
YBX1P1	-0.02
FTH1P7	-0.02
RPL35P5	-0.02
RPL13AP25	-0.02
RPL37AP1	-0.02
AURKB	-0.02
FAUP1	-0.02
MMP9	-0.02
FTLP2	-0.02
RPL7P9	-0.02
GPX1P1	-0.02
CHI3L1	-0.02
MIR621	-0.02
PI3	-0.02
RPL6P27	-0.02
IGFBP2	-0.02
BIRC5	-0.02
TPI1P1	-0.02
FTH1P8	-0.02
RPS27AP16	-0.02
RPS27AP5	-0.02
LOC105369550	-0.02
PLA2G2A	-0.02
NDUFA4P1	-0.02
FTH1P2	-0.02
TMSB4XP8	-0.02
RPL26P19	-0.02
COL1A1	-0.02
COX6A1P2	-0.02
COL3A1	-0.02
RPS27P4	-0.02
SAA1	-0.02
RPL14P1	-0.02
RPS27P3	-0.02
SNRPGP10	-0.02
MT2P1	-0.02
RPS3AP26	-0.02
PBK	-0.02
TOP2A	-0.02
CPXM1	-0.02
MIR3682	-0.02
RPS18P12	-0.03
RPS3AP6	-0.03
SNRPGP2	-0.03
RPS7P1	-0.03
RPS15P4	-0.03
MYBL2	-0.03

RPL41P1	-0.03
RPLP0P9	-0.03
UBE2C	-0.03
RPL18AP3	-0.03
RPL39P3	-0.03
FTLP3	-0.03
RPLP0P6	-0.03

Supplementary Table II. The top 100 relevant genes contributing to the unitary vector along PC2 in the common plot.

Gene	Extended name	Weight in PC1	Phenotypes (GWAS, GenesLikeMe)
SNAP25	Synaptosome Associated Protein 25	1.00	Respiratory, homeostasis/ metabolism, mortality/aging
VSNL1	Visinin Like 1	0.90	Energy expenditure, Alzheimer
STMN2	Stathmin 2	0.90	Dementia
ENC1	Ectodermal-Neural Cortex 1	0.89	Homeostasis/ metabolism
NEFL	Neurofilament Light Chain	0.89	-
SYT1	Synaptotagmin 1	0.88	Energy intake, respiratory, homeostasis/ metabolism, mortality/aging, decreased cell number
RGS4	Regulator Of G Protein Signaling 4	0.86	Homeostasis/ metabolism, craniofacial, mortality/aging
CHN1	Chimerin 1	0.84	Hair color, homeostasis/ metabolism, mortality/aging
GABRA1	Gamma-Aminobutyric Acid Type A Receptor Subunit Alpha1	0.77	Cognitive decline, homeostasis/ metabolism, mortality/aging
GABRG2	Gamma-Aminobutyric Acid Type A Receptor Subunit Gamma2	0.77	Cognitive decline, homeostasis/ metabolism, mortality/aging

Supplementary Table III. The top 10 genes in the PCA ranking for the transition from O to AD **[21]**. Weights in the unitary vector along PC1 are normalized to the highest one. GWAS phenotypes are indicated in the last column.