SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
GENE EXPRESSION
COMPUTATIONAL MODELS
GENETICS OF THE NERVOUS SYSTEM

Received
7 May 2013

Accepted
4 November 2013

Published
19 November 2013

Correspondence and requests for materials should be addressed to J.B.N. (nikas001@umn.edu)

Inflammation and Immune System Activation in Aging: A Mathematical Approach

Jason B. Nikas[1,2]

[1]Department of Neurosurgery, University of Minnesota Medical School, Minneapolis, Minnesota, USA, [2]Masonic Cancer Center, University of Minnesota Medical School, Minneapolis, Minnesota, USA.

Memory and learning declines are consequences of normal aging. Since those functions are associated with the hippocampus, I analyzed the global gene expression data from post-mortem hippocampal tissue of 25 old (age ≥ 60 yrs) and 15 young (age ≤ 45 yrs) cognitively intact human subjects. By employing a rigorous, multi-method bioinformatic approach, I identified 36 genes that were the most significant in terms of differential expression; and by employing mathematical modeling, I demonstrated that 7 of the 36 genes were able to discriminate between the old and young subjects with high accuracy. Remarkably, 90% of the known genes from those 36 most significant genes are associated with either inflammation or immune system activation. This suggests that chronic inflammation and immune system over-activity may underlie the aging process of the human brain, and that potential anti-inflammatory treatments targeting those genes may slow down this process and alleviate its symptoms.

In the absence of any neurodegenerative disease, the aging process of the human brain is inevitably and quintessentially characterized by memory and learning impairments. Unlike in the case of a neurodegenerative disease, normal aging has not been associated with neuronal loss[1–3]. Rather, it has been observed that the impairments induced by normal aging are associated with synaptic remodeling, and that they are more likely to affect functions that are associated with the hippocampus, i.e. several areas of memory and learning[1,2,4].

In order to study the process of human normal aging, this study focuses on the most vulnerable target of that process, namely, the hippocampus. Given the long, gradual course of the normal aging process, I arbitrarily defined the boundaries of the two groups as follows: 1) Old subjects (O) with age ≥ 60 years and 2) Young subjects (Y) with age ≤ 45 years. This 15-yr age gap, I theorized, would accentuate the contrast between the two groups in connection with this otherwise continuous and overlapping process. I analyzed the global gene expression data from post-mortem hippocampal tissue (harvested from the body of the hippocampus at the level of the lateral geniculate nucleus) of 25 old and 15 young cognitively intact human subjects, posted at the Gene Expression Omnibus (GSE11882)[5]. Demographical information pertaining to all 40 subjects is shown in Supplementary Table 1.

Results

Having employed three different and independent methods of statistical significance, namely, ROC curve analysis, fold change, and P-value, I was able to identify 36 genes that were the most significant in terms of differential expression. Fig. 1b depicts the results of K-Means clustering analysis based on the expression of the top 36 most significant genes. All K-Means clustering analysis results (with respect to both the housekeeping genes and the 36 most significant genes) are shown in Supplementary Table 2. As can be seen in both Fig. 1b and Supplementary Table 2, there is a clear separation of the two groups. Fig. 2 depicts the heat map that resulted by plotting the expression of those 36 genes for all 40 subjects (15 young and 25 old). As can be seen by the relative intensities, all of the 36 most significant genes are over-expressed (red color) in the case of the old subjects as compared with the case of the young subjects (blue color). The direction of the differential expression of those 36 genes also appears in Table 1. Moreover, Fig. 3 provides a 3D representation of the differential expression of those 36 genes between the two groups in a surface-contour plot.

Mathematical modeling of aging. Given the aforementioned biovariability in connection with hippocampal gene expression, I wanted to explore whether, via mathematical modeling, I could generate a function that could

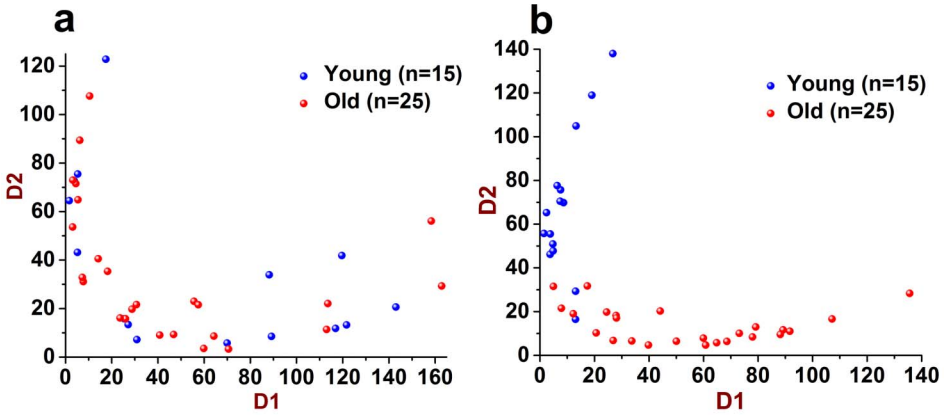

Figure 1 | Results of K-Means clustering analysis. K-Means clustering of the hippocampal tissue gene expression of 25 old subjects (O) and 15 young subjects (Y) with respect to the house-keeping genes (a), and with respect to the 36 most significant genes (b). In (a), in connection with the house-keeping genes, the two groups are inseparable and indistinguishable; whereas in (b), in connection with the 36 most significant genes, the two groups are separated and are clearly distinguishable. D1 is subject distance from the centroid of cluster 1, and D2 is subject distance from the centroid of cluster 2.

classify the 40 subjects with a high accuracy. Such a model would be valuable in future studies of global gene expression analysis of post-mortem hippocampal tissue investigating biological and chronological aging. To that end, I randomly selected approximately 70% of the subjects [11/15 young subjects and 18/25 old subjects] for the development of the function (henceforward referred to as super variable), and I used the remaining subjects (4 young and 7 old) solely for the purpose of validating the super variable. Employing a general methodology that I have previously introduced[6,7], I was able to generate a super variable (function) that, based on the input of 7 genes from the 36 most significant genes, was able to identify/classify accurately all but one of the old subjects {subject # 33 [64 yrs (F)]} [sensitivity = (24/25) = 0.96] and all of the young subjects [specificity = (15/15) = 1.00]. Those overall results of the performance of the F_1 super variable were obtained by combining the results from the development and the validation phases. According to the rank that appears in Table 1, the seven genes that provide the input to the F_1 super variable are: *C4A* (*C4B*), *ADORA3*, *MS4A7*, *BCL6*, *CD44*, *C3AR1*, and *HLA-DRB1*. All of those seven genes are, in terms of biological function, either genes of inflammation or genes of immune system activation. Supplementary Fig. 1 shows the F_1 super variable function in relation to its 7 input gene variables. Fig. 4 and Supplementary Table 3 show the overall results of the F_1 super variable, i.e. the F_1 scores of all 40 subjects used in this study, as well as their respective classification. Fig. 4 and Supplementary Table 3 were created by combining the results from the development phase (the F_1 scores of all 29 subjects that were randomly selected and used exclusively for the development of the model) with the results from the validation phase (the F_1 scores of all 11 subjects that were randomly selected and used exclusively for testing purposes). The results of the F_1 super variable in the development phase are shown in Supplementary Fig. 2 and Supplementary Table 4, whereas the results in the validation phase are shown in Supplementary Fig. 3 and Supplementary Table 5.

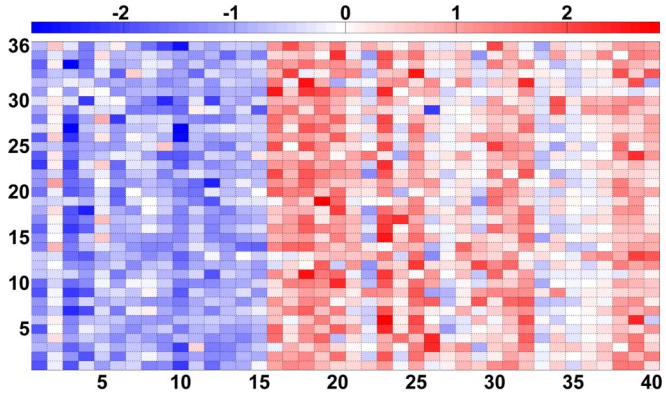

Figure 2 | Heat map of the hippocampal gene expression of all 40 subjects. Heat map of the hippocampal tissue gene expression of 15 young subjects (columns # 1–15) and 25 old subjects (columns # 16–40) with respect to the 36 most significant genes (rows # 1–36). The intensity scale of the standardized expression values ranges from −2.5 (blue: low expression) to +2.5 (red: high expression), with 0 (white) representing the reference intensity value. As can be seen, and based on the group mean expression values, all 36 most significant genes are over-expressed in the case of the old subjects as compared with the young subjects.

It is interesting to note here that, assessing and comparing the performance of the F_1 super variable (Supplementary Table 3) with that of the supervised K-Means clustering (Supplementary Table 2), one can see that the former yielded one misclassification as opposed to four yielded by the latter.

Finally, it should also be noted here that, owing to the constraints of this study, namely, the paucity of healthy, normal human brain tissue samples and respective available data, the F_1 super variable needs to be further validated with a larger, independent cohort.

Biovariability of aging. It has long been observed empirically that aging is not a steady-state, uniformly continuous process; that it is characterized by a relatively wide biovariability; and that biological age may not necessarily coincide with chronological age. The results of my study corroborate those observations. Looking at the expression of the 36 most significant hippocampal genes of all 40 subjects [15 young (columns 1–15) and 25 old (columns 16–40)] in Fig. 2, one notices that four old subjects {# 22, 27, 33, and 35 [80 yrs (M), 83 yrs (M), 64 yrs (F), & 86 yrs (M), respectively]} displayed gene expression patterns that were distinctly closer to those of the young subjects than the patterns of the rest of the old subjects. Conversely, the same observation, albeit in the opposite direction, can be made for one of the young subjects {# 2 [45 yrs (F)]}. The results of K-Means clustering analysis supported numerically those observations (Supplementary Table 2). Moreover, the aforementioned observations about the biovariability of the aging process were also supported by the results of hierarchical clustering analysis performed on the F_1 scores of all 40 subjects (Supplementary Fig. 4).

Table 1 | Top 36 most significantly differentially expressed genes

Rank	Probe Set	Gene Name	Diff. Expr. (O)	ROC AUC	Fold Change	P Value	Key Known Function/Process
1	208451_s_at	*C4A (C4B) LOC100509001*	↑	0.98400	1.22894	2.087E-10	inflammatory response, complement activation, innate immune response
2	206171_at	*ADORA3*	↑	0.96800	1.22659	2.480E-09	inflammatory response, positive regulation of leukocyte migration, histamine secretion by mast cell, signal transduction
3	224358_s_at	*MS4A7*	↑	0.96000	1.23646	7.397E-09	unique expression pattern among hematopoietic cells and nonlymphoid tissues, associated with mature cellular function in the monocytic lineage, and it may be a component of a receptor complex involved in signal transduction
4	215990_s_at	*BCL6*	↑	0.96000	1.13544	3.401E-08	regulation of inflammatory response, regulation of immune response, B cell differentiation, positive regulation of B cell proliferation, regulation of memory T cell differentiation, negative regulation of T-helper 2 cell differentiation, negative regulation of type 2 immune response
5	228532_at	*C1orf162*	↑	0.95733	1.19777	1.617E-07	protein coding, unknown function/process
6	209443_at	*SERPINA5*	↑	0.95733	1.10590	4.544E-08	heparin binding, regulation of blood coagulation, serine protease inhibitor, glycosaminoglycan binding, platelet alpha granule, platelet dense tubular network
7	213566_at	*RNASE6*	↑	0.94933	1.23400	7.506E-08	RNA catabolic process, defense response, ribonuclease activity
8	204489_s_at	*CD44*	↑	0.94933	1.21095	1.042E-08	inflammatory response, wound healing involved in inflammatory response, positive regulation of neutrophil apoptosis, macrophage fusion, neuron projection development
9	232568_at	*MGC24103*	↑	0.94667	1.12816	4.835E-08	unknown function/process
10	209906_at	*C3AR1*	↑	0.94667	1.12143	4.492E-07	complement receptor mediated signaling pathway, inflammatory response, positive regulation of macrophage chemotaxis, positive regulation of neutrophil chemotaxis, elevation of cytosolic calcium ion concentration, signal transduction
11	204912_at	*IL10RA*	↑	0.94667	1.11718	6.043E-07	interleukin-10 receptor activity, response to lipopolysaccharide, signal transducer activity
12	209612_s_at	*ADH1B*	↑	0.94400	1.20914	2.449E-08	metabolic process, ethanol oxidation, reactive oxygen species regulation, noradrenaline & adrenaline & serotonin degradation
13	240578_at	*LOC100507531*	↑	0.94400	1.13432	1.558E-07	unknown function/process
14	212689_s_at	*KDM3A*	↑	0.94400	1.05485	3.431E-08	regulated by peripheral blood monocytes, regulation of transcription, cell differentiation, hormone-mediated signaling pathway, oxidation reduction
15	204787_at	*VSIG4*	↑	0.94133	1.24063	2.204E-07	complement activation, alternative pathway, negative regulation of interleukin-2 production, negative regulation of T cell proliferation, innate immune response
16	208306_x_at	*HLA-DRB1*	↑	0.94133	1.12707	3.108E-07	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II, immune response, signal transduction, MHC class II receptor activity
17	219666_at	*MS4A6A*	↑	0.93867	1.19913	4.561E-07	unique expression pattern among hematopoietic cells and nonlymphoid tissues
18	218084_x_at	*FXYD5*	↑	0.93600	1.08851	2.032E-07	up-regulation of chemokine production, ion transport, actin binding, ion channel activity
19	210184_at	*ITGAX*	↑	0.93600	1.05952	3.555E-07	immune response, IL-8 signaling, integrin signaling, TREM1 signaling, leukocyte migration
20	221698_s_at	*CLEC7A*	↑	0.93333	1.17697	1.946E-07	inflammatory response, T cell activation, innate immune response, positive regulation of phagocytosis, MHC protein binding
21	1560477_a_at	*SAMD11*	↑	0.93333	1.05869	2.586E-08	protein self-association, SAM domain binding, negative regulation of transcription from RNA polymerase II promoter
22	203561_at	*FCGR2A*	↑	0.93067	1.22043	2.682E-06	IgG binding, receptor activity, protein binding
23	225353_s_at	*C1QC*	↑	0.93067	1.20862	3.082E-07	immune response, complement activation (classical pathway), negative regulation of granulocyte differentiation, innate immune response, negative regulation of macrophage differentiation
24	229635_at	UKNOWN GENE	↑	0.92533	1.27241	5.179E-07	unknown function/process

Table 1 | Continued

Rank	Probe Set	Gene Name	Diff. Expr. (O)	ROC AUC	Fold Change	P Value	Key Known Function/Process
25	209823_x_at	*HLA-DQB1*	↑	0.92533	1.16521	5.152E-06	immune response, antigen processing and presentation, MHC class II receptor activity, peptide antigen binding
26	201887_at	*IL13RA1*	↑	0.92533	1.13028	3.689E-07	positive regulation of immunoglobulin production, cell surface receptor linked signaling pathway, positive regulation of B cell proliferation, cytokine receptor activity
27	201137_s_at	*HLA-DPB1*	↑	0.92533	1.12720	1.286E-05	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II, immune response, antigen processing and presentation, MHC class II receptor activity
28	202957_at	*HCLS1*	↑	0.92267	1.22983	2.653E-07	erythrocyte differentiation, positive regulation of cell proliferation, response to hormone stimulus, intracellular signaling pathway
29	227703_s_at	*SYTL4*	↑	0.92267	1.21486	2.891E-06	exocytosis, intracellular protein transport, negative regulation of insulin secretion, transporter activity
30	1559663_at	UNKNOWN GENE	↑	0.92267	1.17895	1.249E-07	unknown function/process
31	210423_s_at	*SLC11A1*	↑	0.92267	1.17552	1.958E-05	inflammatory response, immune response, negative regulation of cytokine production, positive regulation of cytokine production, T cell proliferation involved in immune response, T cell cytokine production, positive regulation of dendritic cell antigen processing and presentation, positive regulation of T-helper 1 type immune response, macrophage activation, positive regulation of phagocytosis
32	1552508_at	*KCNE4*	↑	0.92267	1.17293	1.878E-05	potassium ion transport, voltage-gated potassium channel activity, activation and proliferation of leukocytes
33	237904_at	UNKNOWN GENE	↑	0.92267	1.17094	8.457E-06	unknown function/process
34	209312_x_at	*HLA-DRB1* *HLA-DRB4* *HLA-DRB5*	↑	0.92267	1.13834	1.083E-06	immune response, antigen processing and presentation of peptide or polysaccharide antigen via MHC class II, MHC class II receptor activity
35	208253_at	*SIGLEC8*	↑	0.92000	1.10780	7.415E-07	unique expression in hemopoietic cells, lymphocyte activation signaling, cell adhesion, signal transduction, transmembrane receptor activity
36	226853_at	*BMP2K*	↑	0.92000	1.10059	1.070E-06	protein phosphorylation, protein serine/threonine kinase activity, ATP binding, regulation of bone mineralization (regulated by COX2)

The final 36 most significantly differentially expressed genes between the old and the young subjects, ranked according to their ROC AUC value. The arrows indicate over-expression (↑) or under-expression (↓) of the old subjects (O) as compared with young subjects (Y).

Inflammation and immune system activation in aging. Remarkably, of the 30 known genes out of the 36 most significant genes, 27 were – in terms of function – either genes of inflammation or genes of immune system activation (Table 1). This suggests that – to a large extent, and insofar as it pertains to the hippocampal area of the brain – the dual process of a chronic inflammation and the elicited chronic immune-system response and activity can differentiate between old and young brains with a high accuracy. This is further supported by the fact that the aforementioned seven genes employed by the F_1 super variable, all of which are genes of inflammation or genes of immune system activation, can discriminate between old and young brains with almost a perfect accuracy [sensitivity = (24/25) = 0.96 and specificity = (15/15) = 1.00].

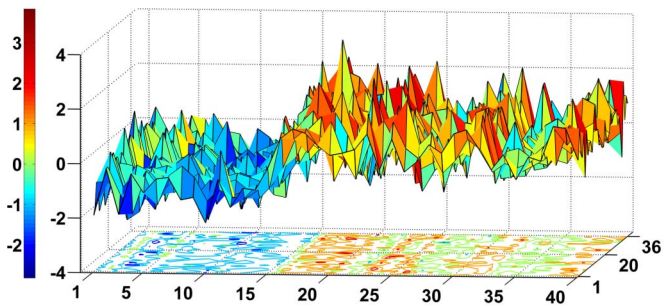

Figure 3 | Surface-contour plot of the hippocampal gene expression of all 40 subjects. Hippocampal tissue gene expression of 15 young subjects (columns # 1–15) (x-axis) and 25 old subjects (columns # 16–40) (x-axis) with respect to the 36 most significant genes (rows # 1–36) (y-axis). The intensity scale of the standardized expression values ranges from −4 (blue: low expression) to +4 (red: high expression) (z-axis). As can be seen, and based on the group mean expression values, all 36 most significant genes are over-expressed in the case of the old subjects as compared with the young subjects.

Discussion

The seven genes [*C4A* (*C4B*), *ADORA3*, *MS4A7*, *BCL6*, *CD44*, *C3AR1*, and *HLA-DRB1*], which are the constituent input variables of the model (F_1 super variable), and all of which are – in terms of function – inflammation or immune system activation genes (Table 1), were all found to be over-expressed in the old subjects compared with the young subjects (Table 1). *C4A* (*C4B*) has been observed to be over-expressed in patients with Huntington disease[8]

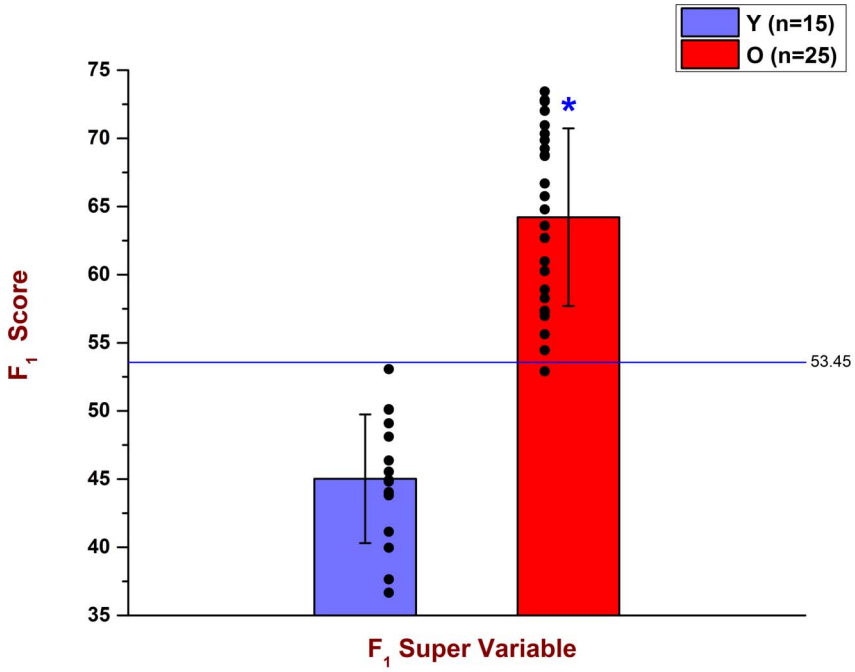

Figure 4 | Overall results of the F_1 super variable (function). The F_1 uses 7 of the 36 most significant genes as its input variables. Using the expression value of those 7 genes for a particular subject, the F_1 yields the F_1 score of that subject; and, based on the determined cut-off score of 53.450, the F_1 classifies that subject as young if the F_1 score is < 53.450 or as old if the F_1 score is ≥ 53.450. As can be seen by the overall performance, the F_1 classified correctly all subjects except one old one [sensitivity = (24/25) = 0.96 and specificity = (15/15) = 1.00]. The mean F_1 score of the Y subjects was 45.028 (top of the blue bar) and their standard deviation (whiskers above or below the top of the blue bar) was 4.721. The mean F_1 score of the O subjects was 64.212 (top of the red bar) and their standard deviation (whiskers above or below the top of the red bar) was 6.514. The significance level was set at $\alpha = 0.001$ (two-tailed), and the probability of significance for the F_1 was $P = 4.18 \times 10^{-12}$ (independent t-Test with T-value = 9.927). The F_1 is parametrically distributed with respect to both groups. The F_1 scores of all 40 subjects are shown in Supplementary Table 3.

and Alzheimer disease[9], in mice with rheumatoid arthritis[10], etc. *ADORA3* has been found to be over-expressed in the hippocampus of patients with Parkinson disease[11], in patients with astrocytomas[12], etc. *MS4A7* has been observed to be over-expressed in mice with rheumatoid arthritis[10]. *BCL6* has been observed to be over-expressed in patients with Huntington disease[13], with ischemic stroke[14], with rheumatoid arthritis[15], with B-cell lymphoma[16], etc. *CD44* has been found to be over-expressed in patients with systemic lupus erythematosus[17], with immune thrombocytopenia[18], with schwannomas[19], with Huntington disease[8], and numerous other diseases and conditions. Over-expression of *C3AR1* has been observed in patients with severe acute respiratory syndrome[20], with asthma[21], etc., while over-expression of *HLA-DRB1* has been observed in patients with multiple sclerosis[22], with rheumatoid arthritis[23], with Duchenne muscular dystrophy[24], etc.

Previous studies using animal models have observed associations between aging and inflammation in connection with the hippocampus, the neocortex, and the cerebellum[25,26]. Using animal models or human subjects with early-stage neurodegenerative diseases, such as Alzheimer, other studies have observed a link between neuroinflammation and deficits in synaptic plasticity, especially long-term potentiation (LTP) in the hippocampus, which is associated with long-term memory consolidation[27–32,2]. The fact that definitive causality cannot be established here notwithstanding – in other words, whether it is the normal aging process that induces inflammation/immune-system-overactivity, or whether the vice versa occurs, or whether another, hitherto unspecified, process engenders the normal aging process, which in turn induces inflammation/immune-system-overactivity, or whether that unspecified process engenders inflammation/immune-system-overactivity, which in turn induces the normal aging process – the results of my study support a direct causal link between the normal aging process and the process of inflammation/immune-system-overactivity. When considered collectively, therefore, the results of my study and all of the above observations from the other aforementioned studies point to a plausible theory on the normal aging process. At some point in time, chronic, low-level inflammation establishes itself and elicits a corresponding chronic immune response and activity. These two conjugate processes ultimately are responsible for a gradual loss of synaptic plasticity, particularly LTP in the hippocampus, accompanied with a minimal neuronal loss[33,34,1–3]. It is this loss of synaptic plasticity – at least in the hippocampus part of the brain – that is associated with the phenotypical changes of normal aging.

The results of my study, in addition to providing evidence for this dual process of chronic, low-level neuroinflammation/immune-system-activation in connection with normal aging, suggest a means of a potential treatment. Regardless of the exact causal sequence of the events, administration of anti-inflammatory drugs/chemicals that can normalize the expression of the aforementioned 27 genes of inflammation/immune-system-activation may decelerate the onset of the aging process, as well as the aging process itself, and mitigate its symptoms by restoring synaptic plasticity throughout the hippocampus and possibly throughout the rest of the brain. Supplementary Table 6 lists all those 27 most significant genes as possible targets for the development of such an anti-inflammatory treatment, along with potential candidate drugs/chemicals that are known (via Ingenuity Pathway Analysis) to interact with those genes.

It is worth noting here that various anti-inflammatory drugs have been used in an effort to slow down the progression of neurodegenerative diseases, such as Alzheimer, with various degrees of success[35,36]. The magnitude of the neuroinflammatory processes in the case of Alzheimer disease or other neurodegenerative diseases, however, cannot be compared to that of the neuroinflammation in the normal aging process; and by virtue of the same argument, the task of

halting neuronal cell loss during the course of a neurodegenerative disease cannot be compared to that of restoring synaptic plasticity during the course of the normal aging process. It would stand to reason, therefore, that anti-inflammatory treatment strategies may be more successful and efficacious than those employed against the progression of neurodegenerative diseases. Finally, I should point out that, based on increased evidence over the last twenty years or so, neuroinflammation seems to be the common denominator of normal aging and neurodegenerative diseases, such as Alzheimer. Understanding the causal circumstances under which a chronic, low-level neuroinflammatory process can transition to a major neuroinflammation conducive to neuronal degeneration and death would be of paramount importance. According to the latest evidence[37], even a single substitution mutation on a single inflammation gene might suffice to trigger that transition in a small percentage of the population.

Methods

Data acquisition. I downloaded the raw intensity microarray data (CEL files) of 25 old subjects (age $\geq$ 60 years) and 15 young subjects (age $\leq$ 45 years) posted at the Gene Expression Omnibus (accession number: GSE11882)[5].

Data processing. I processed the original raw intensity data (CEL files) using the Expression Console software by Affymetrix with the library for the HG-U133 Plus 2.0 microarray chip, and choosing the RMA algorithm (510 (k) FDA approved) with the standard settings.

Statistical methods. I first assessed the quality of the data by examining the expression of all housekeeping genes by all 40 subjects (25 old and 15 young). With regard to the housekeeping genes, there was no statistically significant differential expression between the two groups (Supplementary Table 7). As can be seen by the results of K-Means clustering analysis in Fig. 1a, the two groups cannot be discriminated based on the expression of the housekeeping genes. Having, thus, established the quality of the data, I investigated for any differential expression among all gene variables using three different and independent methods. 1) Using a methodology that I have developed and introduced previously[38–42], I performed ROC curve analysis on all gene variables in order to assess their discriminating power with respect to the two groups (old vs. young), and with respect to this method, I set statistical significance at ROC AUC $\geq$ 0.920. 2) For a given gene variable, I defined fold change (FC) as the mean expression value of the old subjects (O) over the mean expression value of the young subjects (Y), and I set statistical significance at FC $\geq$ 1.10 (if O > Y) or FC $\leq$ 0.91 (if O < Y). 3) I performed independent t-Tests, Mann-Whitney U tests, or Aspin-Welch unequal-variance tests (depending on how a particular gene variable was distributed), and having accounted for all possible comparisons (there are 54,675 probe sets in the Affymetrix HG-U133 Plus 2.0 chip), I set the significance level at $\alpha = 9.15 \times 10^{-7}$. The results according to those three different and independent methods, and prior to the final selection, are shown in Supplementary Table 8. In order to minimize the number of false negatives in the case of the third method[43,44], for the final selection of significant variables, I imposed the condition that if a given gene variable met the significance criteria of all three methods, or those of the first method and those of only one of the other two methods, it would be deemed significant. Excluding multiplicities (different transcripts that corresponded to the same genes), thirty six genes made up the final list of the most significantly differentially expressed genes between the two groups, as assessed by the aforementioned three different and independent methods of statistical significance (Table 1).

In greater detail, to assess statistical significance, I used to assess statistical significance, I used the following three different and independent methods. *1) ROC curve analysis.* I performed ROC curve analysis on all gene variables in order to assess their discriminating power with respect to the two groups (old vs. young), and with respect to this method, I set statistical significance at ROC AUC $\geq$ 0.920. *2) Fold Change.* For all gene variables, I defined fold change (FC) as the mean expression value of the old subjects over the mean expression value of the young subjects, and I set statistical significance at FC $\geq$ 1.10 (if O > Y) or FC $\leq$ 0.91 (if O < Y). *3) P-value.* I used the independent t-Test for parametric gene variables (both normality and homogeneity of variance conditions were met); the Aspin-Welch unequal-variance test (AW) for gene variables that met the normality condition but not the homogeneity of variance condition; and the Mann-Whitney U test (MW) for the non-parametric gene variables, i.e., for those variables that i) the normality condition was not met or ii) the normality and the homogeneity of variance conditions were not met. Taking into account that there are 54,675 probe sets (including those of the housekeeping genes) in the Affymetrix HG-U133 Plus 2.0 chip, and using the Bonferroni correction, I set the significance level for the entire study at $\alpha = 9.15 \times 10^{-7}$. Therefore, in order for any variable to be deemed significant according to the P-value method, the following condition must be met: $P < \alpha$. Regarding the Mann-Whitney U test (MW), since none of the non-parametric variables had any sets of ties (a subject from one group having the same expression value as a subject from the other group), I used the exact probability for all MW tests.

Incorporating the three aforementioned independent methods of statistical significance assessment, and in order to minimize the number of false negatives in the case of the third method[43,44], I set the overall significance criterion as follows: in order for any variable to be included in the final list of the most significant variables, it would have to meet the significance criteria of the first method (ROC AUC $\geq$ 0.920) and those of at least one of the other two methods [FC $\geq$ 1.10 (or FC $\leq$ 0.91) and/or $P < 9.15 \times 10^{-7}$].

Mathematical modeling. Utilizing the final 36 most significant genes, I wanted to explore the possibility of developing – via mathematical modeling – a function that could identify as correctly as possible the age status (O or Y) of an unknown subject based on the expression of any combination of those 36 most significant genes. To that end, I randomly selected approximately 70% of the subjects [11/15 young subjects and 18/25 old subjects] that could be used only for the development phase of such function. In other words, a function could be developed only by the exclusive use of those 29 subjects. The remaining 11 subjects (4 young and 7 old ones) were designated unknown (test) subjects and were used solely for the purpose of validating any promising function generated in the development phase. This split into two fixed sets, whereby one is used only for training and the other only for validation, represents the simplest implementation of K-fold cross validation[45,46]. A function was deemed promising in the development phase only if it exhibited a sensitivity $\geq$ 0.90 and a specificity $\geq$ 0.90 in connection with the 29 subjects of the development phase. Pertaining to the validation phase, and in connection with the 11 unknown subjects, a promising function would have to exhibit the same minimum classification accuracy (a sensitivity $\geq$ 0.90 and a specificity $\geq$ 0.90) in order to be accepted. I was able to generate one such function (F_1 – henceforward also referred to as super variable) that fulfilled all of the aforementioned criteria. Supplementary Fig. 1 shows the equation of F_1 as a function of 7 genes.

The cut-off score of the F_1 was determined by taking into account the results of the following two analyses: 1) calculation of the optimal point on the ROC curve based on the 29 F_1 scores of the 29 subjects used in the development phase [optimal point is defined as the point with the highest sensitivity and the lowest false positive rate (1-specificity)] and 2) calculation of the 99.99% confidence intervals for the mean F_1 scores of the two groups (O and Y) of those 29 subjects and their respective standard deviations. The 99.99% confidence intervals were calculated based on a bootstrap sample size of 100,000. Taking into account the aforementioned ROC optimal point, as well as the relative overlap of MO and MY [MO = LLO – SDO and MY = ULY + SDY] (LLO: the 99.99% confidence lower limit for the mean of the O group; SDO: standard deviation of the O group; ULY: the 99.99% confidence upper limit for the mean of the Y group; SDY: standard deviation of the Y group), the cut-off score of the F_1 super variable was determined to be 53.450. If a subject's F_1 score is < 53.450, then that subject is classified as Y (young); otherwise, if the F_1 score is $\geq$ 53.450, then that subject is classified as O (old). It should be pointed out here that, based on the equation of the F_1 (Supplementary Fig. 1), a given F_1 score is just a numerical value and does not signify age or number of years.

In addition to the main validation method explained above, and in order to further assess the performance of the F_1 super variable, I employed two other and different cross validation methods: 1) a 10-fold cross validation and 2) a leave-one-out cross validation[47]. Both of those methods yielded a misclassification rate of 0.05 and a mean-squared error of 0.05 in connection with the F_1 super variable. The results of those methods, along with the confusion matrices generated by them, are shown in Supplementary Table 9. As can be seen in Supplementary Table 9, each one of those two and different validation methods resulted in a correct classification of all of the young subjects and in a correct classification of all but two of the old subjects.

Computer software. All analyses in this study were carried out with custom software written in MATLAB R2012b. All computer programs in connection with the model were also created using MATLAB R2012b.

1. Morrison, J. H. & Baxter, M. G. The ageing cortical synapse: hallmarks and implications for cognitive decline. *Nat. Rev. Neurosci.* **13**, 240–250 (2012).
2. Bishop, N. A., Lu, T. & Yankner, B. A. Neural mechanisms of ageing and cognitive decline. *Nature* **464**, 529–535 (2010).
3. Pakkenberg, B. *et al*. Aging and the human neocortex. *Exp. Gerontol.* **38**, 95–99 (2003).
4. Benoit, C. E., Rowe, W. B., Menard, C., Sarret, P. & Quirion, R. Genomic and proteomic strategies to identify novel targets potentially involved in learning and memory. *Trends Pharmacol. Sci.* **32**, 43–52 (2011).
5. Berchtold, N. C. *et al*. Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proc. Natl. Acad. Sci. USA* **105**, 15605–15610 (2008).
6. Nikas, J. B., Boylan, K. L., Skubitz, A. P. & Low, W. C. Mathematical prognostic biomarker models for treatment response and survival in epithelial ovarian cancer. *Cancer Inform.* **10**, 233–247 (2011).
7. Nikas, J. B., Low, W. C. & Burgio, P. A. Prognosis of treatment response (pathological complete response) in breast cancer. *Biomark. Insights* **7**, 59–70 (2012).
8. Hodges, A. *et al*. Regional and cellular gene expression changes in human Huntington's disease brain. *Hum. Mol. Genet.* **15**, 965–977 (2006).
9. Rogers, J. *et al*. Complement activation by beta-amyloid in Alzheimer disease. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10016–10020 (1992).

10. Fujikado, N., Saijo, S. & Iwakura, Y. Identification of arthritis-related gene clusters by microarray analysis of two independent mouse models for rheumatoid arthritis. *Arthritis Res. Ther.* **8**, R100 (2006).
11. Xu, K., Bastia, E. & Schwarzschild, M. Therapeutic potential of adenosine A(2A) receptor antagonists in Parkinson's disease. *Pharmacol. Ther.* **105**, 267–310 (2005).
12. Trincavelli, M. L. *et al.* A3 adenosine receptors in human astrocytoma cells: agonist-mediated desensitization, internalization, and down-regulation. *Mol. Pharmacol.* **62**, 1373–1384 (2002).
13. Strand, A. D. *et al.* Gene expression in Huntington's disease skeletal muscle: a potential biomarker. *Hum. Mol. Genet.* **14**, 1863–1876 (2005).
14. Sharp, F. R. *et al.* The future of genomic profiling of neurological diseases using blood. *Arch. Neurol.* **63**, 1529–1536 (2006).
15. Rangel-Moreno, J. *et al.* Inducible bronchus-associated lymphoid tissue (iBALT) in patients with pulmonary complications of rheumatoid arthritis. *J. Clin. Invest.* **116**, 3183–3194 (2006).
16. Pileri, S. A. *et al.* Primary mediastinal B-cell lymphoma: high frequency of BCL-6 mutations and consistent expression of the transcription factors OCT-2, BOB.1, and PU.1 in the absence of immunoglobulins. *Am. J. Pathol.* **162**, 243–253 (2003).
17. Li, Y. *et al.* Phosphorylated ERM is responsible for increased T cell polarization, adhesion, and migration in patients with systemic lupus erythematosus. *J. Immunol.* **178**, 1938–1947 (2007).
18. Jemas, M., Nookaew, I., Wadenvik, H. & Olsson, B. MicroRNA regulate immunological pathways in T-cells in immune thrombocytopenia (ITP). *Blood* **121**, 2095–2098 (2013).
19. Sherman, L. *et al.* CD44 expression is aberrant in benign Schwann cell tumors possessing mutations in the neurofibromatosis type 2, but not type 1, gene. *Cancer Res.* **57**, 4889–4897 (1997).
20. Reghunathan, R. *et al.* Expression profile of immune response genes in patients with Severe Acute Respiratory Syndrome. *BMC Immunol.* **6**, 2.
21. Kohl, J. & Wills-Karp, M. Complement regulates inhalation tolerance at the dendritic cell/T cell interface. *Mol. Immunol.* **44**, 44–56 (2007).
22. Baranzini, S. E. *et al.* Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature* **464**, 1351–1356 (2010).
23. Davidson, A. & Diamond, B. Autoimmune diseases. *N. Engl. J. Med.* **345**, 340–350 (2001).
24. Nogutsi, S. *et al.* cDNA microarray analysis of individual Duchenne muscular dystrophy patients. *Hum. Mol. Genet.* **12**, 595–600 (2003).
25. Terao, A. *et al.* Immune response gene expression increases in the aging murine hippocampus. *J. Neuroimmunol.* **132**, 99–112 (2002).
26. Weindruch, R., Kayo, T., Lee, C. K. & Prolla, T. A. Gene expression profiling of aging using DNA microarrays. *Mech. Ageing Dev.* **123**, 177–193 (2002).
27. Chapman, T. R., Barrientos, R. M., Ahrendsen, J. T., Maier, S. F. & Patterson, S. L. Synaptic correlates of increased cognitive vulnerability with aging: peripheral immune challenge and aging interact to disrupt theta-burst late-phase long-term potentiation in hippocampal area CA1. *J. Neurosci.* **30**, 7598–7603 (2010).
28. Lynch, M. A. Age-related neuroinflammatory changes negatively impact on neuronal function. *Front. Aging Neurosci.* **1**(6), (2010).
29. Rao, J. S., Kellom, M., Kim, H. W. & Rapoport, S. I. Neuroinflammation and synaptic loss. *Neurochem. Res.* **37**, 903–910 (2012).
30. Smith, T. D., Adams, M. M., Gallagher, M., Morrison, J. H. & Rapp, P. R. Circuit-specific alterations in hippocampal synaptophysin immunoreactivity predict spatial learning impairment in aged rats. *J. Neurosci.* **20**, 6587–6593 (2000).
31. Lu, T. *et al.* Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883–891 (2004).
32. Lee, C. K., Weindruch, R. & Prolla, T. A. Gene-expression profile of the ageing brain in mice. *Nat. Genet.* **25**, 294–297 (2000).
33. Rapp, P. R. & Gallagher, M. Preserved neuron number in the hippocampus of aged rats with spatial learning deficits. *Proc. Natl. Acad. Sci. USA* **93**, 9926–9930 (1996).
34. Geinisman, Y. *et al.* Aging, spatial learning, and total synapse number in the rat CA1 stratum radiatum. *Neurobiol. Aging* **25**, 407–416 (2004).
35. Vlad, S. C., Miller, D. R., Kowall, N. W. & Felson, D. T. Protective effects of NSAIDs on the development of Alzheimer disease. *Neurology* **70**, 1672–1677 (2008).
36. Arvanitakis, Z. *et al.* Relation of NSAIDs to incident AD, change in cognitive function, and AD pathology. *Neurology* **70**, 2219–2225 (2008).
37. Jonsson, T. *et al.* Variant of TREM2 associated with the risk of Alzheimer's disease. *N. Engl. J. Med.* **368**, 107–116 (2013).
38. Nikas, J. B., Keene, C. D. & Low, W. C. Comparison of analytical mathematical approaches for identifying key nuclear magnetic resonance spectroscopy biomarkers in the diagnosis and assessment of clinical change of diseases. *J. Comp. Neurol.* **518**, 4091–4112 (2010).
39. Nikas, J. B. & Low, W. C. ROC-supervised principal component analysis in connection with the diagnosis of diseases. *Am. J. Transl. Res.* **3**, 180–196 (2011).
40. Nikas, J. B. & Low, W. C. Application of clustering analyses to the diagnosis of Huntington disease in mice and other diseases with well-defined group boundaries. *Comput. Methods Programs Biomed.* **104**, e133–e147 (2011).
41. Nikas, J. B. & Low, W. C. Linear Discriminant Functions in Connection with the micro-RNA Diagnosis of Colon Cancer. *Cancer Inform.* **11**, 1–14 (2012).
42. Burns, M. B. *et al.* APOBEC3B is an enzymatic source of mutation in breast cancer. *Nature* **494**, 366–370 (2013).
43. Perneger, T. V. What's wrong with Bonferroni adjustments. *BMJ* **316**, 1236–1238 (1998).
44. Feise, R. J. Do multiple outcome measures require p-value adjustment? *BMC Med. Res. Methodol.* **2**, 8 (2002).
45. Efron, B. & Tibshirani, R. *An Introduction to the Bootstrap.* London: Chapman and Hall (1993).
46. Amari, S., Murata, N., Muller, K.-R., Finke, M. & Yang, H. H. Asymptotic Statistical Theory of Overtraining and Cross-Validation. *IEEE Trans. Neural Networks* **8**, 985–996 (1997).
47. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning.* New York: Springer (2001).

Acknowledgments

This study was supported by the Dept. of Neurosurgery and the Masonic Cancer Center at the University of Minnesota. The author would like to thank Douglas Yee and Walter Low for their support.

Author contributions

J.B.N. conceived, designed, and carried out all aspects of this study and wrote and edited the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Nikas, J.B. Inflammation and Immune System Activation in Aging: A Mathematical Approach. *Sci. Rep.* **3**, 3254; DOI:10.1038/srep03254 (2013).