Supporting Information

Berchtold *et al.* 10.1073/pnas.0806883105

SI Text

Tissue Samples. Frozen unfixed tissue was obtained from 55 neuropathologically and neurologically normal controls aged 20–99 years, which were categorized into four age ranges: 20–39, 40–59, 60–79, and 80–99 years (Table S1). Brain regions used in the study were the superior frontal gyrus (SFG), postcentral gyrus (PCG), hippocampus (HC), and entorhinal cortex (EC), and cases were preferentially selected where tissue from at least two brain regions were available from the same individual brain. Cases were 85% Caucasian and 15% African American, balanced across genders and age ranges to avoid error caused by racial skewing. A total of 174 tissue samples were obtained from seven well established National Institute on Aging Alzheimer's Disease brain banks located at the University of California Irvine, Sun Health Research Institute, University of Rochester, Johns Hopkins University, the National Institute of Child Health and Human Development Brain and Tissue Bank for Developmental Disorders at the University of Maryland, University of Pennsylvania, and the University of Southern California. Inclusion and exclusion criteria are indicated in Table S9.

Tissue Dissection, Processing, Gene-Chip Hybridization, and Quality Control. Tissue was obtained from the SFG (crest/superior surface at the genu of the corpus callosum), HC (body of the HC at the level of the lateral geniculate nucleus), EC (crest of the parahippocampal gyrus at the level of the anterior hippocampus), and PCG (crest/superior surface of the PCG where the gyrus meets the superior-most portion of the saggital fissure). Total RNA was extracted from 10–200 mg of frozen, unfixed tissue by using TRIzol reagent (Invitrogen), and purified by using quick spin columns (Qiagen). RNA quality was assessed by using the Agilent BioAnalyzer (Agilent Technologies) and UV spectrophotometer for the presence of sharp peaks and 28S/18S RNA. Each sample was individually hybridized to high-density oligonucleotide gene chips from Affymetrix (human genome Hg-U133 plus 2.0), which measures the expression of >50,000 transcripts and variants, including 38,500 characterized human genes. A total of 174 microarrays were processed at the University of California Irvine DNA and Protein MicroArray Facility, using a robotic system and following manufacturer's recommendations. Briefly, total RNA (10 μg) from each sample was used to generate first-strand cDNA by using a T7-linked-$(dT)_{24}$ primer, followed by *in vitro* transcription using the ENZO BioArray High-Yield RNA transcript labeling kit to generate biotin-labeled cRNA target. Using a robotics system (Biomek FX MicroArray Plex SA System; Beckman Coulter) to optimize consistency in processing and minimize handling variability, each fragmented, biotin-labeled cRNA sample (30 mg) was individually hybridized to an Affymetrix Hg-U133 plus 2.0 chip for 16 h and rotated at 13 rpm at 50°C. The chips were washed and stained on a fluidics station and scanned.

After scanning, CEL files were assessed manually for grid alignment and to ascertain absence of scratches and bubbles. Quality control was assessed by using Affymetrix Expression Console software, following sample and chip quality guidelines from Affymetrix, including recommendation that background levels not exceed 100, and that noise not exceed 4. For all cell files, background levels were <100 (range 35.9- 74.3) and noise was <4 (range 1.72–3.86). Detection of BioB control cRNA is another indicator of chip and sample quality: this control is spiked in at the detection threshold (1.5 pM) and should receive a present detection call in at least 50% of samples, reflecting low noise. In this study, BioB control cRNA received a present call in all samples, indicating good detection levels of even low-abundance transcripts and low noise. Similarly, all other spike-in controls, BioC, BioD and Bio-cre, which are spiked-in at increasing concentrations, had a present detection call in all samples and showed increasing expression across BioC < BioD < Bio-cre, as expected. Most important are 3′ end to 5′ end signal intensity ratios for housekeeping genes such as GAPDH. Because degradation is more likely to occur at the 5′ end of the mRNA molecule that is not protected by a poly(A) tail a high 3′/5′ ratio suggests sample degradation. Affymetrix indicates that housekeeping 3′/5′ ratios should ideally be no higher than 4. GAPDH 3′/5′ ratios in our samples were all <3 (range 0.8–2.6), indicating high RNA integrity for all samples. Finally, the percent of genes with present detection call was within the expected range, with all chips showing percentage present call between 45% and 63%. To minimize technical error, chips were all obtained from the same chip set lot and were run in batches consisting of a random distribution of all samples across tissues, gender, and age ranges.

Microarray Analysis. For low-level processing of Affymetrix GeneChip data, we used two different model-based algorithms (GC-RMA and PLIER) to separately derived two expression values for each probe set represented on each and every array within our experimental design. PLIER and GC-RMA are model-based algorithms that fit a robust linear model to the probe-level data, analyzing the performance of each of the individual probes that make up a specific probe set across each hybridized array in the experiment, applying an error model that assumes that error depends on the individual probes that make up a probe set, rather than on the signal alone. Model-based algorithms such as GC-RMA and PLIER can thus detect abnormally behaving probes more efficiently than nonmodel-based algorithms (such as MAS 5) and generate gene expression values with more reproducible values.

PLIER values were calculated from raw CEL files using Expression Console v1.1 software (Affymetrix) and subsequently exported for analysis in GeneSpring GX software (Agilent Technologies). GC-RMA values, based on the RMA software by Irizarry *et al.* (1, 2), were calculated from raw CEL files within GeneSpring GX. Two different algorithms were used to calculate gene expression values to avoid possible error that can occur with the use of only one method. It is now well understood that the use of different algorithms can result in different calculated gene expression levels for a specific probe set when starting with same originating data scan (e.g., the scanned DAT files). Consequently, depending on the algorithm used to determine the expression levels, a dissimilar set of genes might be identified as being differentially expressed by each method, with varying degrees of resulting overlap found between the two sets. Thus, rather than relying on a single algorithm, a more conservative approach is to use the overlap between multiple algorithms to identify differentially affected genes, which can be used to narrow down a list of new candidate genes by 64–73% (3). We note that while the stringency of the overlapping list reduces false positives, it is also likely that a number of true positives may have also been rejected, whereas the number of false positive and false negatives rejected by the stringency set in the overlapping list is unknown.

In GeneSpring, GC-RMA- and PLIER-derived expression values were processed in parallel, using default settings for

per-chip and per-gene normalization, followed by log-transformation of the geometric mean of expression values. To generate a list of reliably detectable probe sets, probe sets were first filtered on Flag detection calls to remove probe sets that were absent on >50% of the chips for a given region (assignment of detection calls was based on probe-pair intensities for which one probe is a perfect match of the reference sequence and the other is a mismatch probe for which the 13th base (of the 25 oligonucleotide reference sequence is changed). Filtering on Flag detection call at the 50% present level reduces probe sets with unreliable signal and very effectively reduces the incidence of false positives (4). This step reduced the probe sets size from >50,000 probe sets to ≈34,000 for each brain region, and is referred to as the reliable probe set list. Unsupervised clustering of groups and cases for expression across the brain was performed by using Pearson correlation (average linkage algorithm with separation ratio 1, minimum distance 0.001) on the reliable probe set list and cases that had values for multiple brain regions. Subsequently, the reliable probe set list was analyzed in parallel for significant probe sets based on GC-RMA and PLIER determined expression values, with the significance threshold set at $P < 0.01$ using group sizes balanced for region/age/sex, as indicated in Table S2. To increase the internal consistency of the data, only cases where samples were available from at least three brain regions from the same case were used. Variance measures were based on the cross-gene error model in GeneSpring, which estimates measurements of precision by combining variability of gene expression data based on replicate samples. This error model assumes that variability between replicates is comparable for all genes with similarly determined expression levels and uses this information to calculate standard deviation, which is then used in subsequent statistical calculations. Before fitting the error model, the genes are ordered by their control strengths. A median standard deviation and median control strength are calculated for each nonoverlapping set of 11 genes.

To generate a final stringent list of robust probe sets, significant lists based on GC-RMA ($P < 0.01$) and PLIER ($P < 0.01$ derived data were compared, and the probe sets identified as significant when using both methods were selected as robust genes for further analysis. The genes were then filtered on confidence in GeneSpring, using a confidence level of 95% with the Benjamini and Hochberg false discovery rate applied for multiple testing correction, confirming that 90–98% of the probe sets identified in the stringent list met this additional stringency level.

Lists were data-mined by using bioinformatics software in GeneSpring to identify pathways and functional categories that were significantly represented in our gene lists from each brain region. The P values were calculated by using a hypergeometric distribution where P represents the probability of particular mapping arising by chance, given the numbers of genes constituting a list of differentially expressed genes, relative to the set of all genes in the category. The relative weight of each represented pathway is indicated, making highly significant pathways and functions immediately apparent.

PCR Validation. Aging-dependent changes in immune and inflammation-related genes were validated by RT-PCR using hippocampal RNA a subset of the young (20–59 years, $n = 10$) and aged (60–99 years, $n = 10$) cases used in the microarray analysis, focusing on several key factors including complement component C3, CD14, TLR 2, TLR 4, TLR 7, and TOLLIP (Fig. S4). Quantitative RT-PCR was undertaken at the University of Rochester Functional Genomics Center, using the TaqMan Gene Expression Assays (Applied Biosystems) and a GUSB Endogenous control assay. Each TaqMan Gene Expression Assay is preformulated consisting of two unlabeled PCR primers at a final concentration of 900 nM and one FAM dye-labeled TaqMan MGB of 250 nM final concentration. Sequences of the primers and probe are proprietary information. For each sample, 3.0 μg total RNA was reverse-transcribed by using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Two microliters of a 1:5 dilution for cDNA was combined with TaqMan Universal PCR Master Mix No AmpErase UNG (Applied Biosystems) and the Taqman Gene Expression Assay in a 10-μl reaction set up by the CAS-1200 liquid handling system. The real-time RT-PCR amplifications were run on an ABI PRISM 7900 HT Sequence Detection System (Applied Biosystems). Universal thermal cycling conditions were as follows: 10 min at 95°C, 40 cycles of denaturation at 95°C for 15 s, and annealing and extension at 60°C for 1 min. Amplification efficiencies were close to 100% for all assays according to analyses of a number of different dilutions of the cDNA. Calculations were done assuming that 1 delta Ct equals a 2-fold difference in expression. Statistical significance was determined by t tests for each probe by using Statview software.

1. Irizarry RA, *et al.* (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31:e15.
2. Irizarry RA, *et al.* (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249–264.
3. Millenaar FF, *et al.* (2006) How to decide? Different methods of calculating gene expression from short oligonucleotide array data will give different results. *BMC Bioinformatics* 7:137.
4. McClintick JN, Edenberg HJ (2006) Effects of filtering by Present call on analysis of microarray experiments. *BMC Bioinformatics* 7:49.

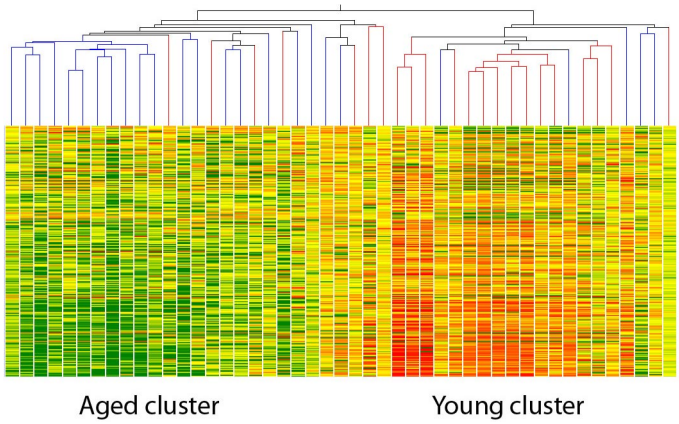

Fig. S1. Unsupervised hierarchical clustering of cases with multiple brain regions based on Pearson correlation of probe sets with present call across >50% of chips (≈33,000 probe sets) reveals that >80% of cases separate into distinct clusters of young (20–59 years) or old (60–99 years) cases. Four of 22 young cases and 5 of 25 aged cases showed closer similarity to the nonrespective age group.

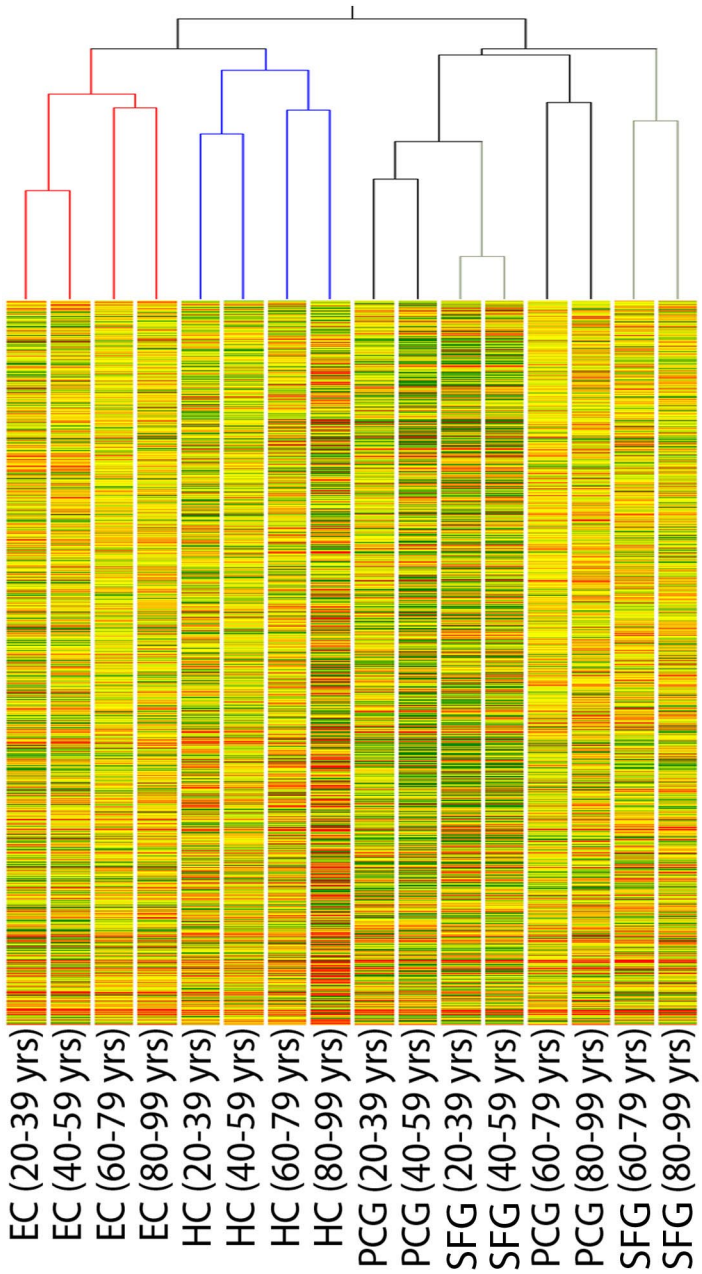

Fig. S2. Unsupervised hierarchical clustering suggests that each brain region has a distinct signature gene expression change with aging across the four age groups (20–39, 40–59, 60–79, and 80–99 years). Age was a more dominant clustering factor than region for the superior frontal gyrus (SFG) and postcentral gyrus (PCG), whereas the hippocampus (HC) and entorhinal cortex (EC) clustered first by region than by age. In all regions, the youngest age groups (20–39 and 40–59 year) clustered and the two older cohorts (60–79 and 80–99 years) clustered.

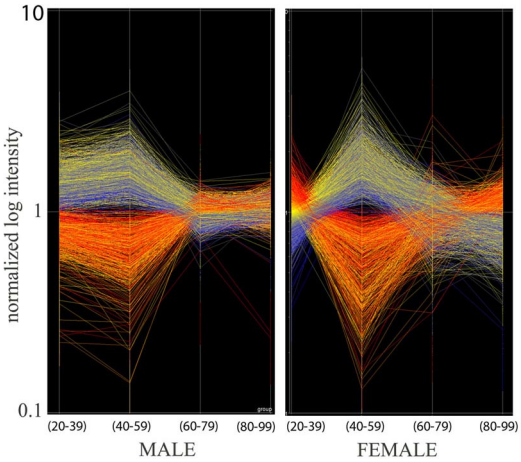

Fig. S3. Genes show sexually dimorphic patterns of expression across aging. In females, a large number of genes undergo a striking transient change in expression levels in the fourth and fifth decades of life, in the postcentral gyrus (PCG). In the male PCG, these genes do not follow a similar expression pattern. Each line represents the average expression level of one probe set at each of the four age groups (20–39, 40–59, 60–79, and 80–99 years).

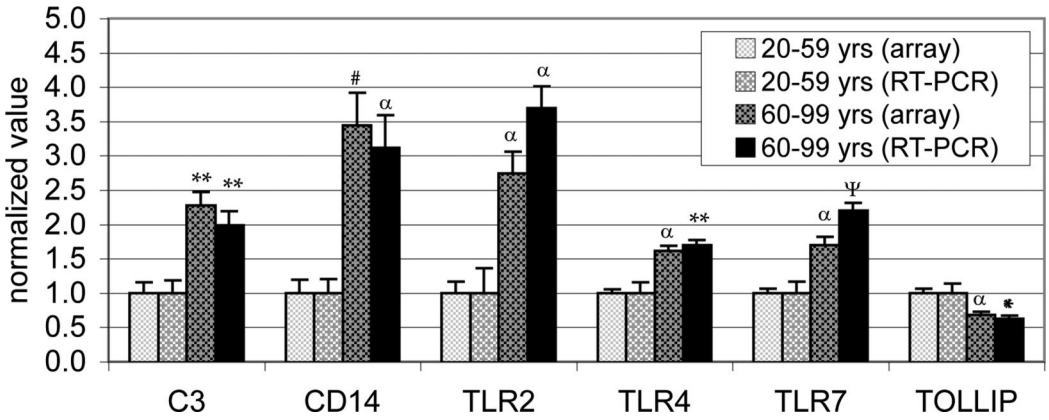

Fig. S4. RT-PCR confirmed aging-dependent changes in immune- and inflammation-related genes in the HC. Both RT-PCR and microarray revealed up-regulation of complement component C3, as well as CD14 and TLRs (TLR2, TLR4, and TLR 7) and down-regulation of TOLLIP, an inhibitor of the Toll-like signaling pathway, in aging. Values were compared between young (20–59 years) vs. aged (60–99 years) and were normalized to respective young expression levels. Similar aging-depending changes were detected by microarray and RT-PCR. Averages ± SEM: *, $P < 0.05$, **, $P < 0.01$, #, $P < 0.001$, Ψ, $P < 0.005$, α, $P < 0.0001$, compared with respective young group.

Other Supporting Information Files

Table 1
Table 2
Table 3
Table 4
Table 5
Table 6
Table 7
Table 8
Table 9