Supplemental Material

MicroRNA, mRNA, and Protein Expression Link Development and Aging in Human and Macaque Brain

Supplemental Tables

Table S1. Sample characteristics (human).

The table lists age, sex and death cause of the subjects, postmortem intervals in hours (PMI), RNA integrity values of RNA samples (RIN - an indicator of RNA quality calculated by Agilent® Bioanalyser 2100). The microarray experiments were carried out in two batches. Two individuals are included in both batches as technical replicates. The column with the "Exp" header indicates the experiments that each sample was used in: **a-** mRNA microarray, **b-** miRNA-sequencing, **c-** protein sequencing. Note that technical and demographic characteristics are independent of subject age: Spearman correlations of age with sex, PMI or RIN are -0.02, -0.10, 0.07, respectively (p>0.5). Two individuals, H5 and H19, overlap with a previously published dataset on brain development (the DFPLC dataset in (Somel et al. 2009)). Abbreviations: ASCVD: arteriosclerotic cardiovascular disease; HASCVD: hypertensive arteriosclerotic cardiovascular disease. For 5 elder individuals we have no detailed information on exact birth dates besides the year of birth; for these, age in years has been used.

ID	Year	Day	Sex	Batch	Exp	RIN	PMI	Cause of death	Overlap
H1	0	2	m	1, 2	a,b,c	8	3	complications of prematurity	3
H2	0	4	m	1	a,b,c	8.8	5	congenital heart defect	5
H3	0	34	m	1	a,b,c	7.9	7	idiopathic pulmonary hemorrhage	7
H4	0	204	m	1	a,b,c	8.4	6	sudden infant death syndrome	6
H5	8	2	m	1	a,b,c	8.3	5	cardiac arrhythmia	5
H6	13	360	m	1	a,b,c	8.3	13	hanging	13
H7	25	152	m	1	a,b,c	9.2	19	asthma	19
H8	53	112	m	1	a,b,c	8.3	17	ASCVD	17
H9	66	0	m	1	a,b,c	8.6	10	ruptured abdominal aneurysm aorta	10
H10	80	0	m	1	a,b,c	8.6	7	ventricular fibrillation	7
H11	88	0	m	1	a,b,c	7.7	7	euthanasia	7
H12	98	0	m	1, 2	a,b,c	7.3	9	cardiac tamponade due to bleeding from aorta fissure	9
H13	0	19	f	2	a	7.1	14	pneumonia associated with meconium aspiration	14
H14	0	94	m	2	a	7.7	12	bronchopneumania	12
H15	1	78	m	2	a	7.6	19	asthma	19
H16	2	57	f	2	a	7.5	21	acute myocarditis	21
H17	4	170	f	2	a	7.7	21	lymphocytic myocarditis	21

H18	16	271	m	2	a	9.1	15	accident, drowning	15
H19	22	334	m	2	a	7.3	4	ASCVD	4
H20	39	74	m	2	a	7.9	12	HASCVD	12
H21	58	34	m	2	a	8.4	9	HASCVD	9
H22	78	222	f	2	a	8	3	natural	3
H23	90	0	f	2	a	7.8	4	natural	4

Table S2. Sample characteristics (macaque).

The table lists age and sex of the subjects and RNA integrity values of RNA samples (RIN - an indicator of RNA quality calculated by Agilent Bioanalyser 2100). Postmortem intervals for all individuals were < 20 minutes. The microarray experiments were carried out in two batches. Two individuals are included in both batches as technical replicates The column with the "Exp" header indicates the experiments each sample was used in: **a-** mRNA microarray, **b-** miRNA-sequencing.

ID	Year	Day	Sex	Batch	Exp	RIN
M1	0	16	m	1, 2	a,b	9.1
M2	0	20	m	1	a,b	9.9
M3	0	153	m	1	a,b	9.8
M4	0	207	m	1	a,b	9.7
M5	0	310	m	1	a,b	9.5
M6	2	9	m	1	a,b	9.0
M7	4	27	m	1	a,b	9.0
M8	9	104	m	1	a,b	9.0*
M9	20	91	m	1	a,b	9.0
M10	22	74	m	1	a,b	8.5*
M11	28	0	f	1, 2	a,b	7.8
M12	26	28	m	1	a,b	8.8
M13	0	22	m	2	a	9.3
M14	0	151	m	2	a	9.1
M15	0	179	m	2	a	9.5
M16	0	237	m	2	a	9.0
M17	1	84	m	2	a	9.2
M18	1	242	m	2	a	8.6
M19	3	40	m	2	a	8.7
M20	8	16	m	2	a	8.7
M21	15	3	m	2	a	8.1
M22	21	8	m	2	a	8.7
M23	25	166	f	2	a	8.2
M24	25	0	f	2	a	8.8

* RIN could not be calculated by the machine, but was estimated from the gel pictures.

Table S3. Proportions of transition genes in mRNA, miRNA and protein datasets.

The table shows the proportion of genes identified as "transition genes" (Methods and represented in Figure 3A). These are genes that show significant change with age (FDR<0.1%), have non-linear trajectories in multiple regression tests and show significant transition points (F-test p<0.05). For the protein dataset and a previously published human dataset (Somel et al. 2009), we chose genes reliably detected and changing significantly with age among the 4,084 age-related genes in the human cortex dataset. The transition points and their significance is estimated separately using log transformed and linear age scales, which capture developmental and post-developmental changes, respectively.

Dataset	Species	% of age-related genes[a]	n. age-related genes	% transition genes (log-age) [b]	% transition genes (linear-age) [b]	% union (log or linear)
mRNA	Human	32.9	4084	60.6	66.8	88.7
mRNA*	Human	21.6	3723	56.6	68.1	85.0
mRNA	Macaque	20.8	2005	27.5	35.5	56.5
miRNA	Human	30.8	115	50.4	38.3	67.0
miRNA	Macaque	20.6	69	29.4	23.5	40.6
Protein	Human	19.7	895	12.5	17.0	25.2

[a] Proportion of genes showing significant change with age among all expressed genes in a dataset.

[b] Proportion of genes showing a significant transition point and having a non-linear trajectory, among all genes showing significant change with age.

* An independent dataset of postnatal mRNA changes in human cortex (Somel et al. 2009).

Table S4. miRNA changes with age in human and mouse.

The table shows the comparison of two studies on mouse post-natal cerebral with miRNA expression changes during human development measured in our study. The signs indicate increase (+) or decrease (-) with age during development, based on Table 1 in (Dogini et al. 2008), visually assessed from Figure 2 in (Smirnova et al. 2005), or calculated using linear regression in the present human cortex dataset. NA indicates that the miRNA is not present in the human miRNA dataset. (?) indicates an ambiguous case, where the hsa-mir-140-5p was decreasing while hsa-mir-140-3p was increasing.

	mir-124a	mir-125a	mir-125b	mir-130	mir-140	mir-205	mir-9	mir-181a	mir-199a	mir-301
Human cortex*	-	-	-	-	?	NA	-	-	-	-
Dogini et al.	-	-	-	-	-	-	-	-	-	-

	mir-9	mir-23	mir-29	mir-125	mir-128
Human cortex*	-	+	+	-	-
Smirnova et al.	-	+	+	-	-

* Present study

Table S5. miRNA-target pairs involved in human cortex development and aging.

The table shows the miRNA-target gene group pairs chosen based on two criteria: (1) the miRNA showing target enrichment in a co-expressed gene group (at HT $p<0.05$), (2) the miRNA showing significantly more negative correlation (at $r<-0.75$) with its targets in that group (at binomial test $p<0.05$), compared to correlations between miRNAs *without* target enrichment in that group and their targets in that group. The regulators are identified based on correlations during development (0-20 years) or aging (20-98 years), separately. The target genes listed in the table are those negatively correlated with their regulator (at $r<-0.75$). The columns labeled "Macaque" and "Protein" indicate whether the miRNA-target gene group pairs are supported (+) or not (-), either by the rhesus macaque data, or by human protein data, respectively. In these cases, we only require a higher number of negatively correlated targets per miRNA, compared to the background (i.e. negatively correlated targets per non-enriched miRNAs), irrespective of statistical significance. NA: the miRNA is not detected in macaque.

Period	Gene group	miRNA	Maca-que	Protein	Target genes
Development	1	hsa-mir-212	-	+	SRGAP2, JARID1B, HN1, AFF2, MAF, C7orf60, SEMA4G, OSBPL8, DYRK2, LEMD3, DPYSL3, C9orf30, TMEFF1, ZNF362, SOS1, SPAST, KIAA1211, SOX5, CTDSPL2, SOX4, SOX11, PCGF3
Development	1	hsa-mir-29a	+	+	NID1, SRGAP2, JARID1B, DGKD, KIRREL, CLDN1, HN1, MAP2K6, BACH2, CACNG4, C11orf57, COMMD2, PTP4A1, KLHDC10, AMMECR1L, AFF2, C7orf60, ZNF346, RALGPS1, BCORL1, UBTD2, NKIRAS2, SERBP1, DCX, RCOR1, ROD1, ROBO1, AP3M1, MMP2, PCDHA13, PCDHAC1, PCDHA10, PCDHAC2, DTX4, NCOA3, KIAA2022, IL17RD, TET1, VASH1, EDC3, ANKRD13B, ISLR2, FAM123B, TRAF4, KLHL8, ZBTB10, COL4A2, COL4A1, CSNK1G1, EML4, ZNF362, MMP24, NAV2, NKAIN1, SPAST, KIAA0895, GNG2, WDR40A, FRMD4A, LCORL, MYCN, ZNF518B, USP42, DOT1L, PXDN, TUBB2B, POLR1D, SDK1, PCGF3
Development	1	hsa-mir-29b	+	+	NID1, SRGAP2, JARID1B, NAV1, DGKD, KIRREL, CLDN1, HN1, MAP2K6, BACH2, CACNG4, C11orf57, COMMD2, PTP4A1, KLHDC10, AMMECR1L, AFF2, C7orf60, ZNF346, RALGPS1, BCORL1, UBTD2, LSM11, NKIRAS2, SERBP1, DCX, RCOR1, ROD1, ROBO1, AP3M1, MMP2, PCDHA13, PCDHAC1, PCDHA10, PCDHAC2, DTX4, NCOA3, KIAA2022, IL17RD, TET1, VASH1, EDC3, ANKRD13B, ISLR2, FAM123B, TRAF4, KLHL8, ZBTB10, COL4A2, COL4A1, CSNK1G1, EML4, ZNF362, MMP24, NAV2, NKAIN1, SPAST, KIAA0895, GNG2, WDR40A, FRMD4A,

					HMGCS1, LCORL, MYCN, USP6NL, ZNF518B, USP42, DOT1L, PXDN, TUBB2B, POLR1D, SDK1, PCGF3
Development	1	hsa-mir-29c	+	+	NID1, SRGAP2, JARID1B, NAV1, DGKD, KIRREL, CLDN1, HN1, MAP2K6, BACH2, CACNG4, C11orf57, VANGL1, COMMD2, PTP4A1, KLHDC10, AMMECR1L, AFF2, C7orf60, ZNF346, RALGPS1, BCORL1, UBTD2, LSM11, NKIRAS2, SERBP1, DCX, RCOR1, ROD1, ROBO1, AP3M1, MMP2, PCDHA13, PCDHAC1, PCDHA10, PCDHAC2, DTX4, NCOA3, KIAA2022, IL17RD, TET1, VASH1, EDC3, ANKRD13B, ISLR2, FAM123B, TRAF4, KLHL8, ZBTB10, COL4A2, COL4A1, CSNK1G1, EML4, ZNF362, MMP24, NAV2, NKAIN1, SPAST, KIAA0895, GNG2, WDR40A, FRMD4A, HMGCS1, LCORL, MYCN, USP6NL, ZNF518B, USP42, DOT1L, PXDN, TUBB2B, POLR1D, SDK1, PCGF3
Development	1	hsa-mir-338-5p	+	+	SRGAP2, JARID1B, KIAA0408, C6orf174, MARCKS, BACH2, RND3, ACVR2A, MED13, PTP4A1, AMMECR1L, ATXN7L1, BCORL1, USP15, TSHZ3, RNF139, AP3M1, B4GALT5, NNAT, CSNK1G1, ZNF362, SPAST, FRMD4A, MYCN, AFAP1, MEIS2, WHSC1, BTG3
Development	1	hsa-mir-499-5p	+	+	TRIM67, SRGAP2, MARCKS, MED13, PTP4A1, DYRK2, LRCH2, ZSWIM5, RRAGB, CSNK1G1, FAM110B, EML4, NAV2, SOX5, EIF4G2, WDR40A
Development	4	hsa-mir-34a	+	+	GREM2, TNRC4, CAMSAP1, WDR68, TANC2, NAV3, LRRC7, SLC16A2, E2F5, BMP3, FKBP1B, PTPRD, GRM7
Development	4	hsa-mir-34c-5p	+	+	GREM2, TNRC4, CAMSAP1, WDR68, TANC2, NAV3, LRRC7, SLC16A2, E2F5, BMP3, EVI5L, FOXJ2, PTPRD, GRM7
Development	6	hsa-mir-212	+	+	HNRNPU, RC3H1, APAF1, ZNF644, ZCCHC11, CCDC88A, CBFA2T2, SLAIN2
Development	6	hsa-mir-22	NA	-	LAMC1, HNRNPH1, MYST4, CHD9, MURC, TET2, C5orf24, EPC1
Development	6	hsa-mir-222	-	+	APAF1, REV1, MED1, TET2, RFX7, SLAIN2, ATAD2B
Development	6	hsa-mir-23b	+	+	HNRNPU, ENAH, ZNF292, ZC3H12C, CTTNBP2NL, APAF1, ATP11C, CFDP1, BTAF1, MTSS1, ADNP, PBRM1, QSER1, SETD2, UBA6, TNPO1
Development	6	hsa-mir-29a	+	+	ENAH, LAMC1, DYNLT1, REV3L, SH3GLB1, SH3PXD2B, MARK3, BRWD3, C5orf24, NUP160, CCDC88A, C16orf88, EPC1, RFX7, RNF122, ZNF532, LRP6, ATAD2B, SCML2
Development	6	hsa-mir-29b	+	+	ENAH, LAMC1, DYNLT1, REV3L, SH3GLB1, SH3PXD2B, MARK3, BRWD3, C5orf24, NUP160, CCDC88A, C16orf88, EPC1, RFX7, RNF122, ZNF532, LRP6, ATAD2B, SCML2
Development	6	hsa-mir-29c	+	+	ENAH, LAMC1, DYNLT1, REV3L, SH3GLB1, SH3PXD2B, MARK3, BRWD3, C5orf24, NUP160, CCDC88A, C16orf88, EPC1, RFX7, RNF122, ZNF532, LRP6, ATAD2B, SCML2
Development	6	hsa-mir-433	-	+	HNRNPU, ENAH, LAMC1, MARK3, MED1, CHD9, TIA1, PBRM1, CEP135
Aging	4	hsa-mir-1271	+	+	ARL4C, UCK2, RAP2B, ODZ4, L1CAM, LRRC7, GRIA1, TSPAN14, EPHA3, SEPT11, MYO16, CUGBP2, BASP1, GRM7
Aging	4	hsa-mir-33b	+	-	PTGFRN, TANC2, ADRA2A, LHX6, REEP1, ROBO2, GRIN3A, PRICKLE2, CALN1,

					BMP3, RNF165, NRIP1
Aging	4	hsa-mir-34a	+	+	GREM2, TNRC4, DPYSL4, CAMSAP1, WDR68, TANC2, NAV3, NRXN2, SLC16A2, B4GALT2, CALN1, RGMB, EVI5L, PNOC, FKBP1B, FOXJ2, RTN4RL1
Aging	6	hsa-mir-181c	+	+	ENAH, MLL3, BPTF, ZNF292, ATP11C, CHD9, QSER1, BIRC6, CEP135, TNPO1, LRP6
Aging	6	hsa-mir-222	-	+	SH3PXD2B, MED1, TET2, CCDC88A, ATAD2B
Aging	6	hsa-mir-433	+	+	MED1, CHD9, TIA1, PBRM1, CEP135

Table S6. miRNA-target pairs overlapping with experimentally verified target sets.

The table lists miRNA-target gene pairs predicted in the current study (based on the target site enrichment and correlation tests; shown in Table S5), which overlap with four datasets of experimentally verified miRNA targets: Tarbase (Papadopoulos et al. 2009), Mirwalk (http://www.ma.uni-heidelberg.de/apps/zmf/mirwalk/contact.html), Khan et al. (Khan et al. 2009) and Baek et al. (Baek et al. 2008). Left columns show a particular miRNA and its targets found in both the current study and the study/database in question. Right columns indicate whether the relationship was estimated in development or aging (using 20 years of age as cutoff). The significance of each overlap ("*p*-value") was estimated by a randomization test (Methods). Expected: the number of target-gene pairs expected to overlap between the current study and a particular database by chance.

Baek et al 2008	
miR-181 TNPO1 ATP11C ENAH	**Aging**
p-value=0.002 Expected=0	
Khan et al 2009	
miR-34a E2F5	**Development**
miR-34c-5p EVI5L	**Development**
miR-34a FOXJ2 DPYSL4	**Aging**
p-value<0.001 Expected =0	
Tarbase	
miR-29c COL4A1 COL4A2 LAMC1	**Development**
p-value<0.001 Expected =0	
Mirwalk	
miR-29a COL4A1 COL4A2 LAMC1	**Development**
miR-29b COL4A1	**Development**

COL4A2	
NKIRAS2	
SERBP1	
LAMC1	
miR-29c	
COL4A1	**Development**
COL4A2	
LAMC1	
miR-34a	
GRM7	**Development**
NAV3	
miR-34a	
FOXJ2	
EVI5L	**Aging**
NAV3	
B4GALT2	
p-value=0.003	
Expected =3.5	

Table S7. Gene Ontology categories enriched among gene groups.

The table shows results from an enrichment analysis of gene groups among Gene Ontology (GO) "biological process" categories (Ashburner et al. 2000), conducted using the func_hyper program (Prüfer et al. 2007). Each gene group is compared to all 4,084 age-related genes. For each gene group, GO categories are chosen based on enrichment in genes in that group at HT $p<0.05$. The table shows five significantly enriched categories with the largest numbers of genes. The global *p*-value column indicates the probability that the overall distribution of genes in test group among all GO "biological process" categories is non-random compared to the background group, estimated by 1,000 permutations of genes across gene groups (Prüfer et al. 2007). Note that not all co-expressed groups are significant in the global enrichment test.

Gene group	**Global *p***	**Enriched Gene Ontology categories**				
1	0.001	Regulation of RNA metabolic process	Regulation of transcription, DNA-dependent	Nervous system development	Positive regulation of transcription, DNA-dependent	Regulation of cell cycle
2	0.004	Protein transport	Intracellular protein transport	Cellular homeostasis	Electron transport chain	Neurotransmitter transport
3	0.005	Oxidation reduction	Response to wounding	Cytoskeleton-dependent intracellular transport	Fatty acid biosynthetic process	Amino acid catabolic process
4	0.038	Cell communication	Cell motion	Cell-cell adhesion	Potassium ion transport	Neuron development
5	0.004	Transport	Regulation of biosynthetic process	Oxidative phosphorylation	Electron transport chain	ATP biosynthetic process
6	0.003	RNA metabolic process	Regulation of transcription, DNA-dependent	mRNA processing	Negative regulation of gene expression	Chromatin modification

7	0.139	Response to external stimulus	Integrin-mediated signaling pathway	Microtubule-based movement	Cell-matrix adhesion	mRNA transport
8	0.005	Cell projection organization	Potassium ion transport	Calcium ion transport	GABA signaling pathway	Cell-substrate adhesion

Table S8. KEGG pathways enriched among gene groups.

The table shows results from an enrichment analysis of gene groups among KEGG pathway categories (Kanehisa et al. 2008), conducted using in-house R code following the same scheme as the func_hyper program (Prüfer et al. 2007). Each gene group is compared to all 4,084 age-related genes. For each gene group, KEGG categories are chosen based on enrichment in genes in that group at HT $p<0.05$. The table shows five enriched categories with the largest number of genes. The global *p*-value column indicates the probability that, for each gene group, the distribution of genes among all KEGG categories is non-random, estimated by 1,000 permutations of genes across gene groups (Methods).

Gene group	**Global *p***	**Enriched KEGG pathways**				
1	0.134	Pathways in cancer	Natural killer cell mediated cytotoxicity	TGF-beta signaling pathway	Fcε RI signaling pathway	Systemic lupus erythematosus
2	0.005	Metabolic pathways	Oxidative phosphorylation	Huntington's disease	Alzheimer's disease	Parkinson's disease
3	0.011	Valine, leucine and isoleucine degradation	Metabolism of xenobiotics by cytochrome P450	Sphingolipid metabolism	Propanoate metabolism	Histidine metabolism
4	0.011	Axon guidance	Neuroactive ligand-receptor interaction	Cell adhesion molecules (CAMs)	Prostate cancer	Long-term depression

5	0.082	Metabolic pathways	Oxidative phosphorylation	Huntington's disease	Parkinson's disease	Alzheimer's disease
6	0.799	Ribosome	Ubiquitin mediated proteolysis	-	-	-
7	0.128	Small cell lung cancer	ECM-receptor interaction	Adipocytokine signaling pathway	Selenoamino acid metabolism	Methionine metabolism
8	0.17	Neuroactive ligand-receptor interaction	MAPK signaling pathway	Calcium signaling pathway	Tight junction	Heparan sulfate biosynthesis

Supplemental Figures

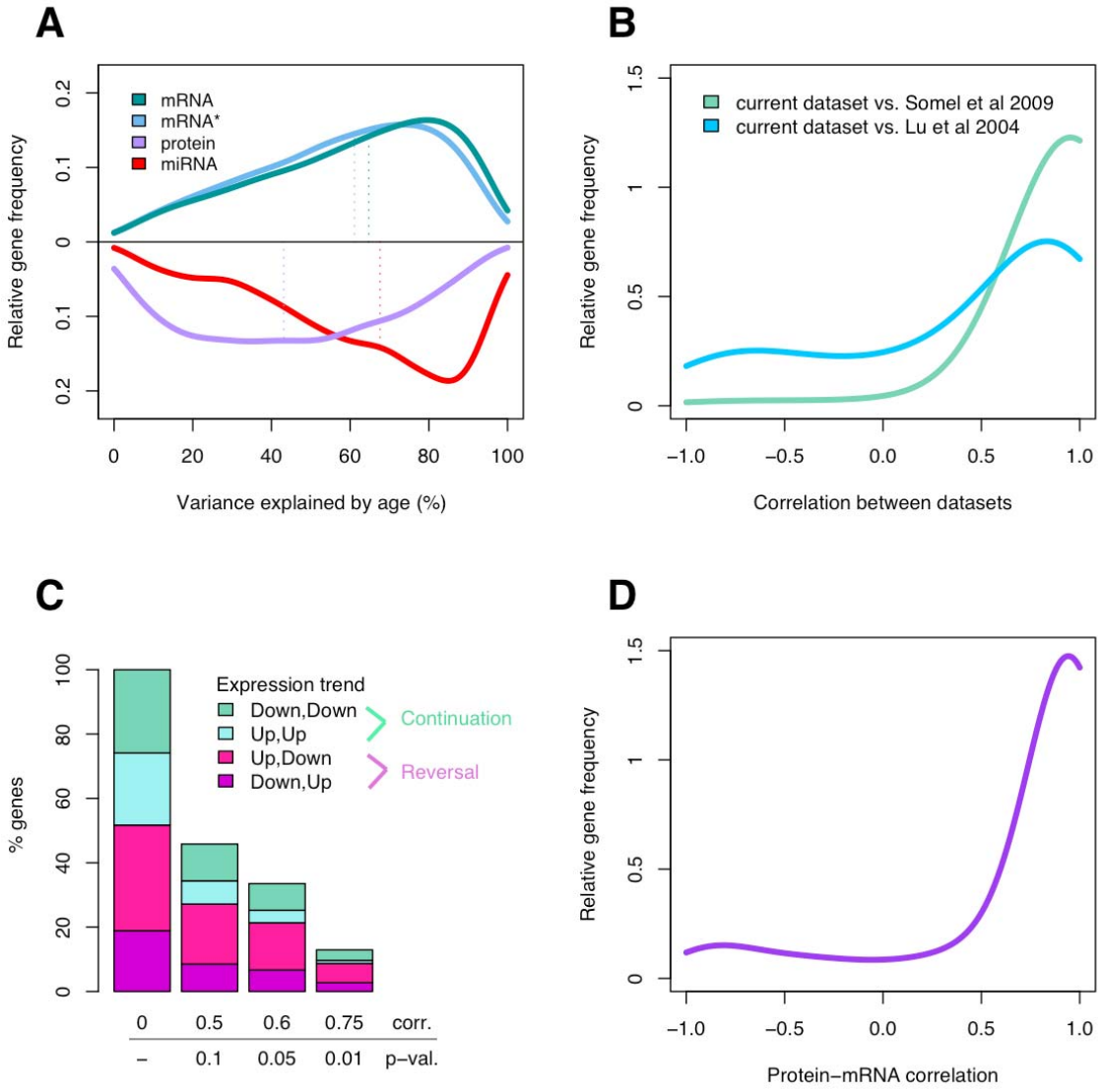


Figure S1. Age effect on expression and correlation between datasets.

(A) Amount of variance explained by age in the mRNA, miRNA and protein datasets, calculated using polynomial regression models. The distributions represent age effect in batch 1 (dark blue) and batch 2 (light blue) samples of the mRNA dataset and age effect in miRNA (red) or protein expression (purple) among the same 12 batch 1 individuals. **(B)** The distribution of Pearson correlation coefficients age-related genes across genes in the human cortex dataset used in this study and reliably measured in a second (Somel et al. 2009) or third (Lu et al. 2004) human cortical gene expression dataset (3723 and 2038

genes, respectively). The y-axis represents gene frequency as Gaussian kernel density. The correlations are based on spline curves describing expression change with age (degrees of freedom=3) estimated for each gene in each pair of datasets, calculated for the common age window between two datasets. Note that all three experiments involve approximately the same brain region (prefrontal cortex), all use different Affymetrix platforms. Two of 44 individuals (with ages 8 and 22.9) in the second dataset (the DFPLC dataset in (Somel et al. 2009)) overlap with those in the current dataset. **(C)** Expression change trends between development and aging. 4,084 age-related genes in the human cortex dataset are classified as being up- or down-regulated during development or aging (before and after 20 years of age, respectively). Expression change trends are estimated based on Pearson correlation between age and expression levels during either period. Genes are sorted at different correlation cutoffs, as indicated on the x-axis. The "Up-Up" and "Down-Down" labels represent up-regulation and down-regulation in both development and aging, respectively. The "Down-Up" and "Up-Down" labels represent down-regulation in development and up-regulation in aging, and *vice versa*. **(D)** Correlation between mRNA and protein expression changes in the human prefrontal cortex. The 901 genes shown in the histogram are chosen as expressed at reliable levels in the protein dataset (total peptide reads≥20) and showing significant expression change with age in the mRNA dataset. The curve represents correlations based on interpolated points from cubic spline curves, describing expression change with age in protein and mRNA.

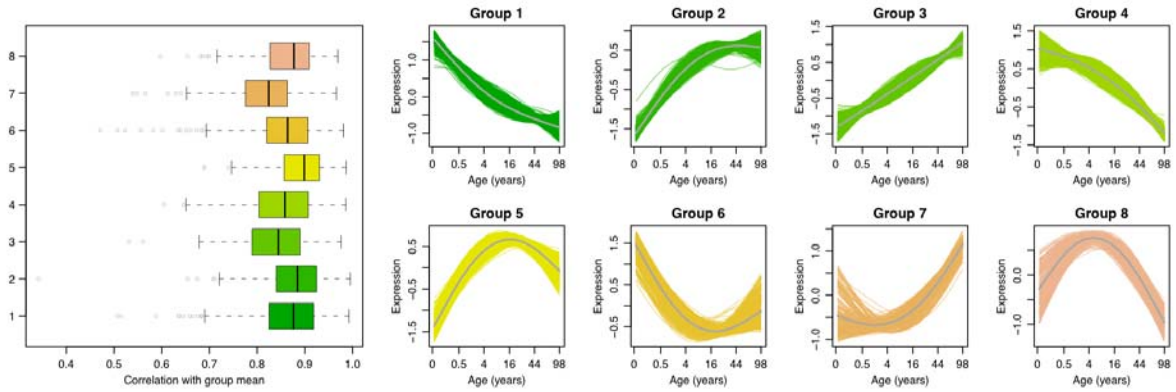


Figure S2. Correlation among genes within gene groups.

Left panel: distribution of Pearson correlation coefficients between each gene groups' members' expression profiles and gene groups' average. Other panels: Cubic spline curves representing standardized expression profiles of each member gene of each group. The grey line indicates the mean. The y-axis shows the standardized expression levels of genes, the x-axis shows age of individuals, in $(age)^{1/4}$ scale.

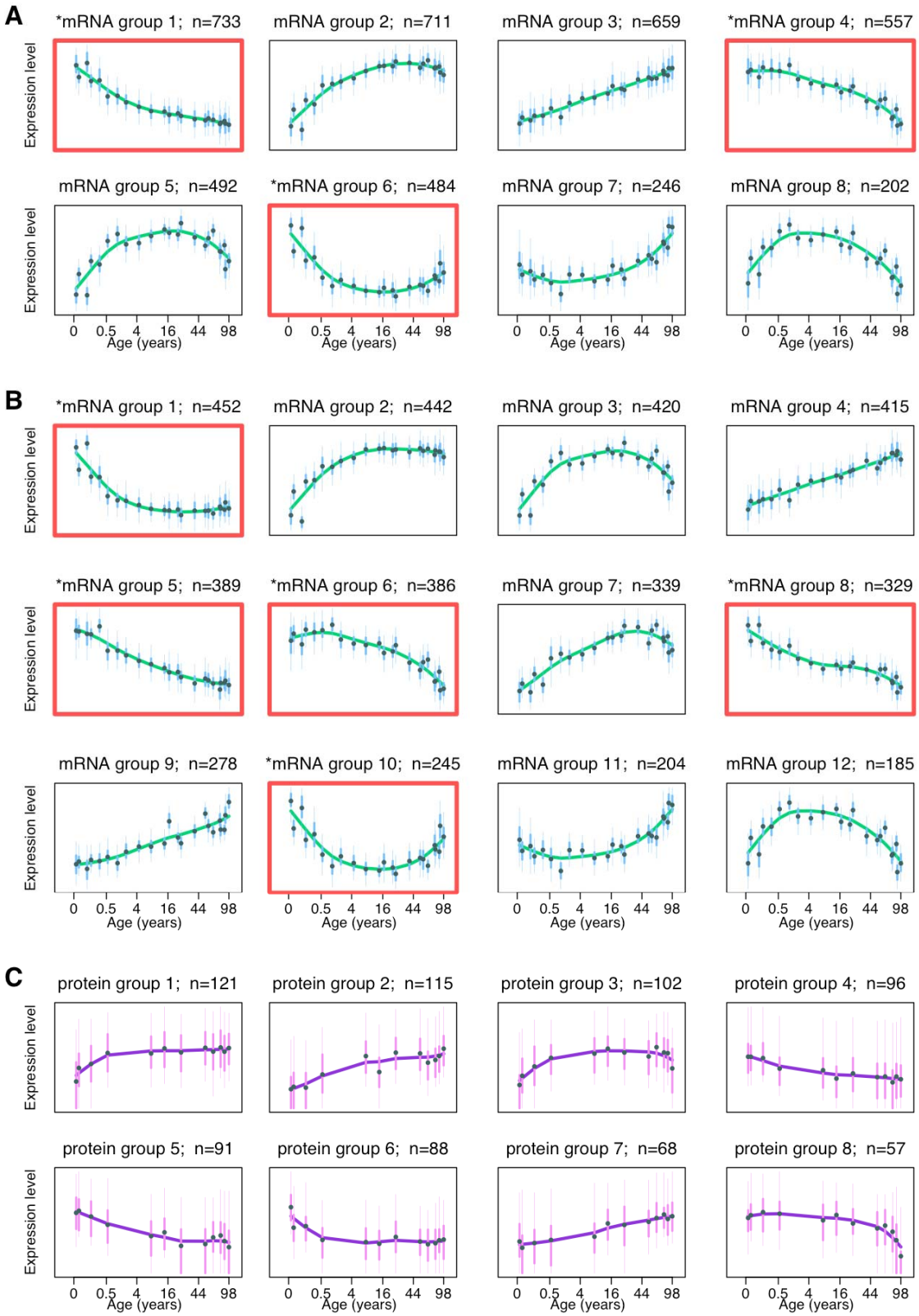

A
*mRNA group 1; n=733
mRNA group 2; n=711
mRNA group 3; n=659
*mRNA group 4; n=557
mRNA group 5; n=492
*mRNA group 6; n=484
mRNA group 7; n=246
mRNA group 8; n=202
B
*mRNA group 1; n=452
mRNA group 2; n=442
mRNA group 3; n=420
mRNA group 4; n=415
*mRNA group 5; n=389
*mRNA group 6; n=386
mRNA group 7; n=339
*mRNA group 8; n=329
mRNA group 9; n=278
*mRNA group 10; n=245
mRNA group 11; n=204
mRNA group 12; n=185
C
protein group 1; n=121
protein group 2; n=115
protein group 3; n=102
protein group 4; n=96
protein group 5; n=91
protein group 6; n=88
protein group 7; n=68
protein group 8; n=57
Expression level
0 0.5 4 16 44 98
Age (years)

Figure S3. Co-expressed mRNA and protein groups.

The graph shows the mean expression levels among co-expressed mRNA and protein groups in the human cortex, plotted against age. **(A-B)** 4,084 age-related genes are clustered into either 8 **(A)** or 12 groups **(B)** using k-means clustering, based on their normalized expression profiles among 23 individuals in the human cortex mRNA dataset. Notably, when both groupings are tested for enrichment in miRNA targets (Methods), we find target enrichment in similar expression change patterns (groups showing an excess of miRNA targets are indicated by the red frames). **(C)** 8 k-means clusters of 738 age-related proteins in the human cortex protein expression dataset, based on 12 individuals. Age-related proteins were chosen based on the age-test (at F-test $p<0.01$, FDR<15%). In all panels, the points represent the mean standardized (z-transformed) expression level of all genes in a group, per individual. The vertical bars indicate the expression variation in the 25%-75% quantile range. The fitted lines correspond to the spline curves and summarize mean expression change with age within the gene group.

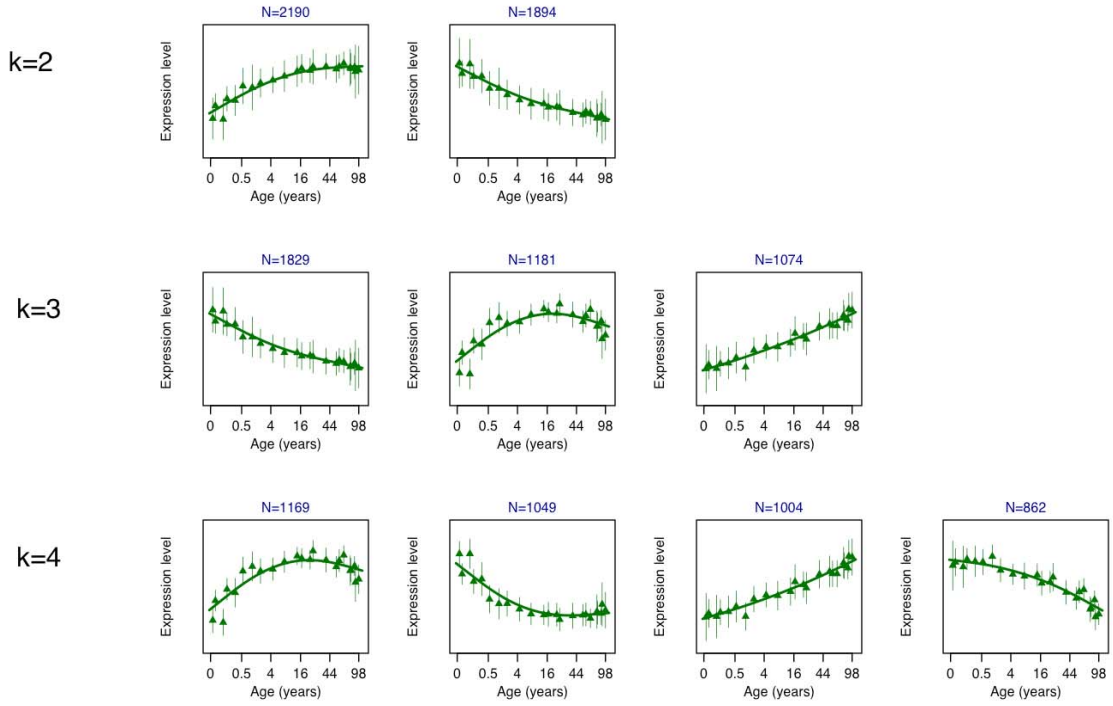


Figure S4. Gene groups showing reversal.

The figure shows results k-means clustering of age-related genes, using 2, 3 and 4 groups. The y-axis shows the mean standardized expression levels of genes in a group, the x-axis shows age of individuals, in $(age)^{1/4}$ scale. Vertical bars show the standard deviation across genes in a group. Notice that reversal between expression changes in early and late life phases can already be detected at the 3 group level.

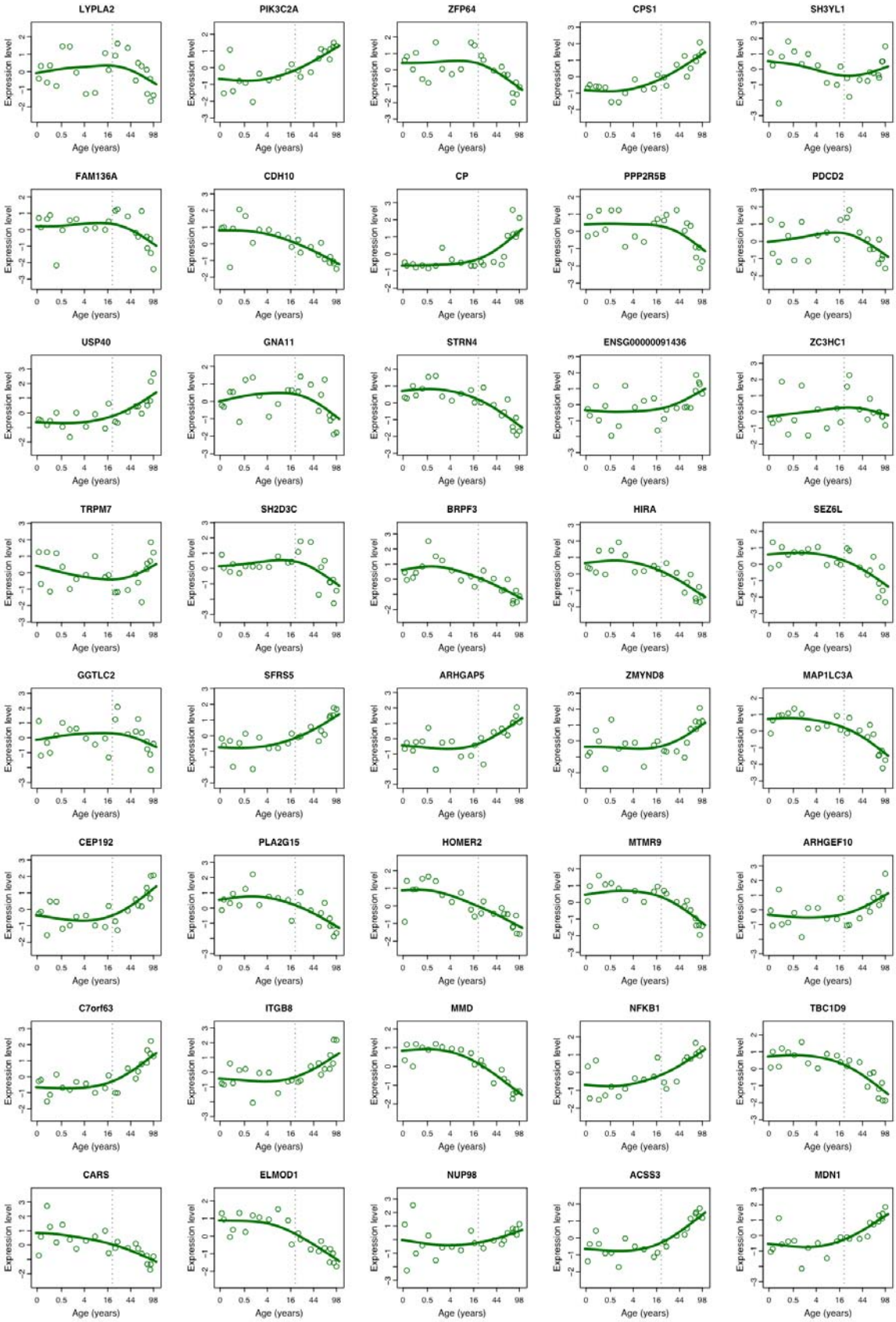

LYPLA2
PIK3C2A
ZFP64
CPS1
SH3YL1
FAM136A
CDH10
CP
PPP2R5B
PDCD2
USP40
GNA11
STRN4
ENSG00000091436
ZC3HC1
TRPM7
SH2D3C
BRPF3
HIRA
SEZ6L
GGTLC2
SFRS5
ARHGAP5
ZMYND8
MAP1LC3A
CEP192
PLA2G15
HOMER2
MTMR9
ARHGEF10
C7orf63
ITGB8
MMD
NFKB1
TBC1D9
CARS
ELMOD1
NUP98
ACSS3
MDN1
Expression level
Age (years)
0
0.5
4
16
44
98

Figure S5. Examples of mRNA expression changes during aging.

The plots show mRNA expression trajectories of genes with strong correlation (Pearson $|r|>0.8$, $p<0.01$) between expression and age after 20 years of age and with no correlation ($|r|<0.2$, $p>0.5$) before 20 years of age, in the human cortex. The 40 examples are randomly chosen among the total of 157 such "aging only" genes found in the human cortex dataset. The y-axis shows the standardized expression levels of genes, the x-axis shows age of individuals, in $(age)^{1/4}$ scale. Dots represent expression levels of individuals, lines represent cubic spline curves. Notably, despite being chosen for showing no statistically significant correlation during development, many genes show visible tendencies of change with age before 20 years of age (shown as the dotted vertical line), although blurred by large inter-individual variance. Thus, genes that show expression changes exclusive to aging are rare. Furthermore, we test whether these genes show significant functional differentiation from gene groups with similar profiles but with clear developmental change trajectories. Specifically, we compare "aging only" genes that decrease during aging, with genes in group 4 as defined in the main text (Figure 2). We similarly compare "aging only" genes that increase during aging with genes in group 7. In both cases, we find no significant differentiation between the two groups using the FUNC tool (global enrichment $p> 0.05$).

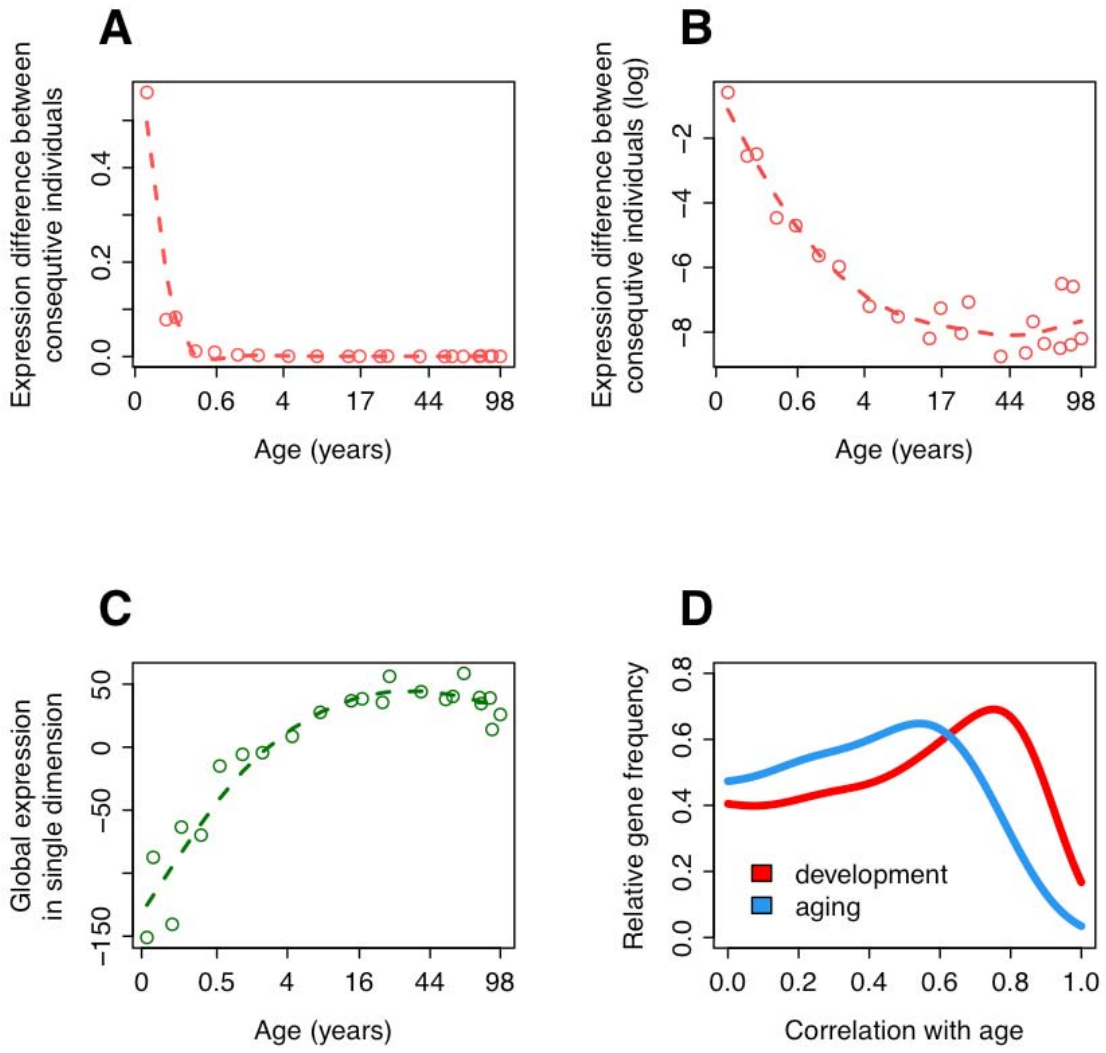


Figure S6. Decreasing magnitude of expression changes in the brain during lifetime. **(A)** Absolute difference in expression levels between individuals of consecutive age, normalized by their age difference. The expression differences are calculated using expression levels standardized to mean=0 and standard deviation=1, per gene. **(B)** Same as Panel A, but the y-axis is log transformed to display changes in late life. **(C)** Global expression change in a single dimension. We employed multidimensional scaling (using the "isoMDS" function in the R "MASS" package) to calculate a single value summarizing expression variation across all expressed genes (following (Somel et al. 2009)). The measure is analogous to the first principle component of a PCA. Note the large magnitude of change in early life, which subsides by adolescence. In Panels **A-C**, x-

axes show individuals' age, in $(age)^{1/4}$ scale. **(D)** Correlation of expression changes with age during development (<20 years of age) and aging (>20) in the human prefrontal cortex. The x-axis shows Pearson correlation coefficient between expression and individual age. The y-axis shows the relative frequency of correlation values across age-related genes, calculated using Gaussian kernel density estimation (using the "density" function in R). Note the lower correlations during aging relative to during development.

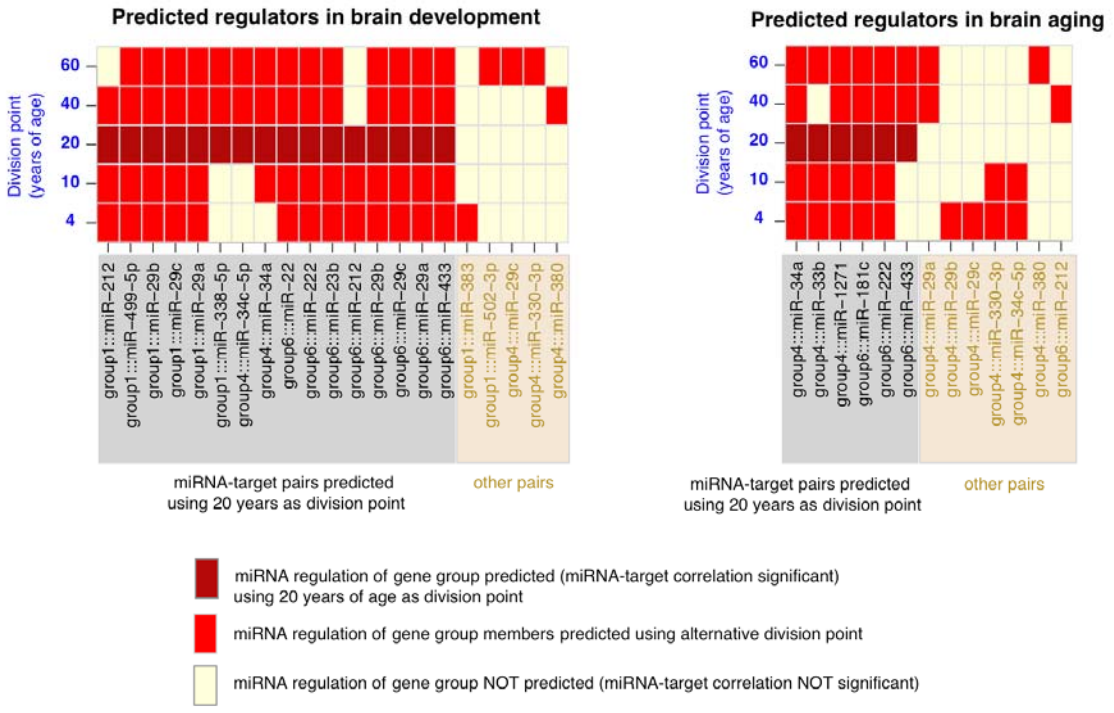


Figure S7. miRNA-target regulation prediction across lifespan.

Our regulator prediction uses 20 years of age as division point between development and aging. Here we tested the influence of the division point choice on regulator estimation. The figure shows the putative regulator miRNA-gene group pairs, estimated during development and aging (left and right panels, respectively), using alternative division points (4, 10, 40, and 60 years of age). Regulator miRNA-gene group pairs are estimated using the distribution of correlation coefficients across individuals in that age range. E.g. at division point 40, we calculate correlations across all individuals younger than 40 ("development") and older than 40 ("aging"). Red: a miRNA-gene group pair is supported by excess of negative correlations, using a certain division point. Yellow: a pair is not supported. Dark red: miRNA-gene group pairs estimated using 20 years as division point and reported in the main text and Table S5 (also highlighted grey). Pairs highlighted light brown are estimated using other division points. Notice that putative miRNA-gene group pairs estimated using 20 years tend to be supported by other division points.

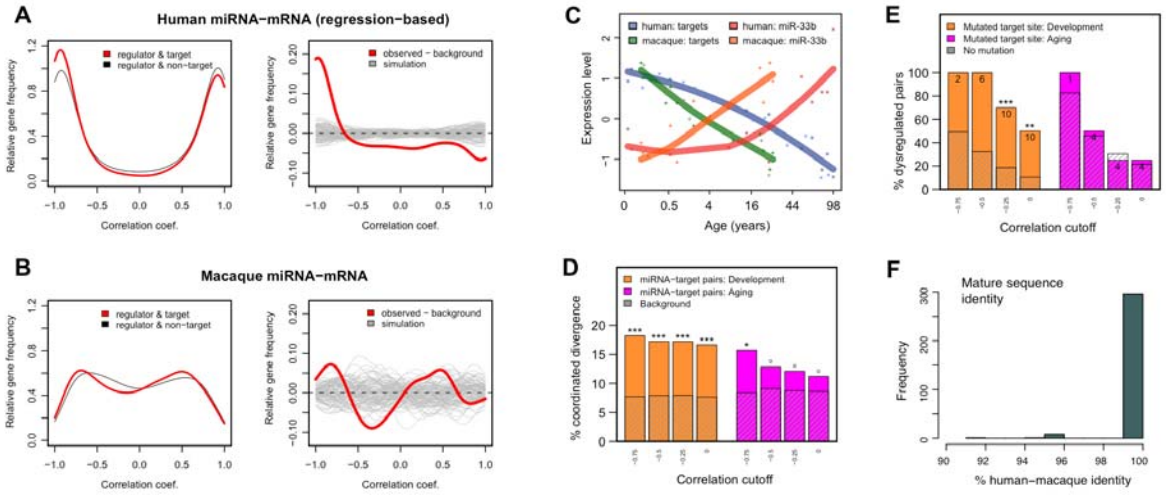


Figure S8. Testing miRNA regulation of mRNA changes.

(A) Excess of negative correlations between miRNA-target pairs in the human dataset. In contrast to Figure 4A, correlations here are based on mRNA and miRNA/TF levels interpolated using spline models of expression change with age (degrees of freedom = 3). This step ensures that the miRNA-mRNA correlations are largely independent of individual variation unrelated to age. Left panel: The density distribution of Pearson correlation coefficients between miRNA and genes, both showing significant expression change with age. The colored lines (red) indicate correlations between regulators and their targets. The grey lines indicate correlations between regulators and non-targets (genes not supposed to be targeted by those regulators) and represent the background. Right panel: The difference between the density distributions of regulator-target correlations and regulator - non-target correlations. The grey lines represent 100 simulation results, where a selection of random regulator-target pairs, the same number as regulator-target pairs, were randomly chosen and compared to the background. **(B)** Excess of negative correlations between miRNA-target pairs in the rhesus macaque dataset. The two panels are drawn following panel (A), using correlations based on original expression data in the rhesus macaque dataset (*i.e.* without interpolation). **(C)** An example of coordinated divergence between human and rhesus macaque expression profiles for miR-33b and its 12 target genes falling into co-expressed group 4. The y-axis

represents the z-transformed miRNA and mean mRNA expression levels, and the solid lines show spline curves of expression changes with age. **(D)** Testing whether miRNA expression divergence between human and rhesus macaque is reflected in their target gene expression. The y-axis shows the proportion of putative regulatory miRNA-target pairs (Table S5) that show coordinated human-macaque expression divergence, compared to random miRNA-target pairs (Methods). The statistics are calculated separately for development and for aging. Here and in panel E asterisks indicate significance based on HT. ***: $p<0.001$, **: $p<0.01$, *: $p<0.05$, ^{o}: $p<0.10$. **(E)** Testing whether a substitution in the miRNA binding site in rhesus macaque causes dysregulation, *i.e.* loss of correlation between miRNA and target gene expression in rhesus macaque. The y-axis shows the proportion of putative regulatory miRNA-target pairs that contain a mutation in the rhesus macaque binding site and lack miRNA-target correlation in macaques (*i.e.* that are dysregulated). The proportion is compared to putative regulatory miRNA-target pairs that do *not* contain a mutation, but are dysregulated in macaque. The numbers above the bars indicate the number of mutated pairs observed at that correlation cutoff. **(F)** Sequence identity between human and macaque mature miRNA sequences detected in both species (n=306).

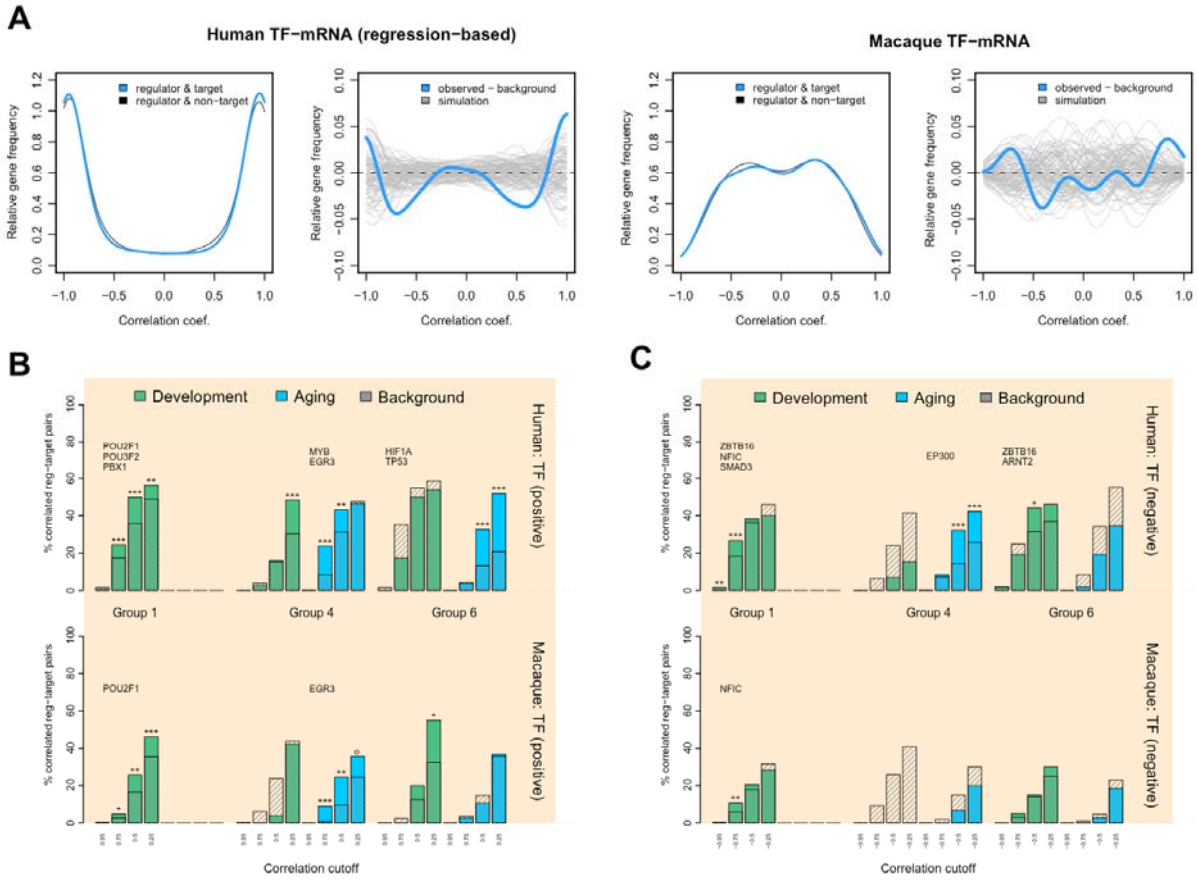
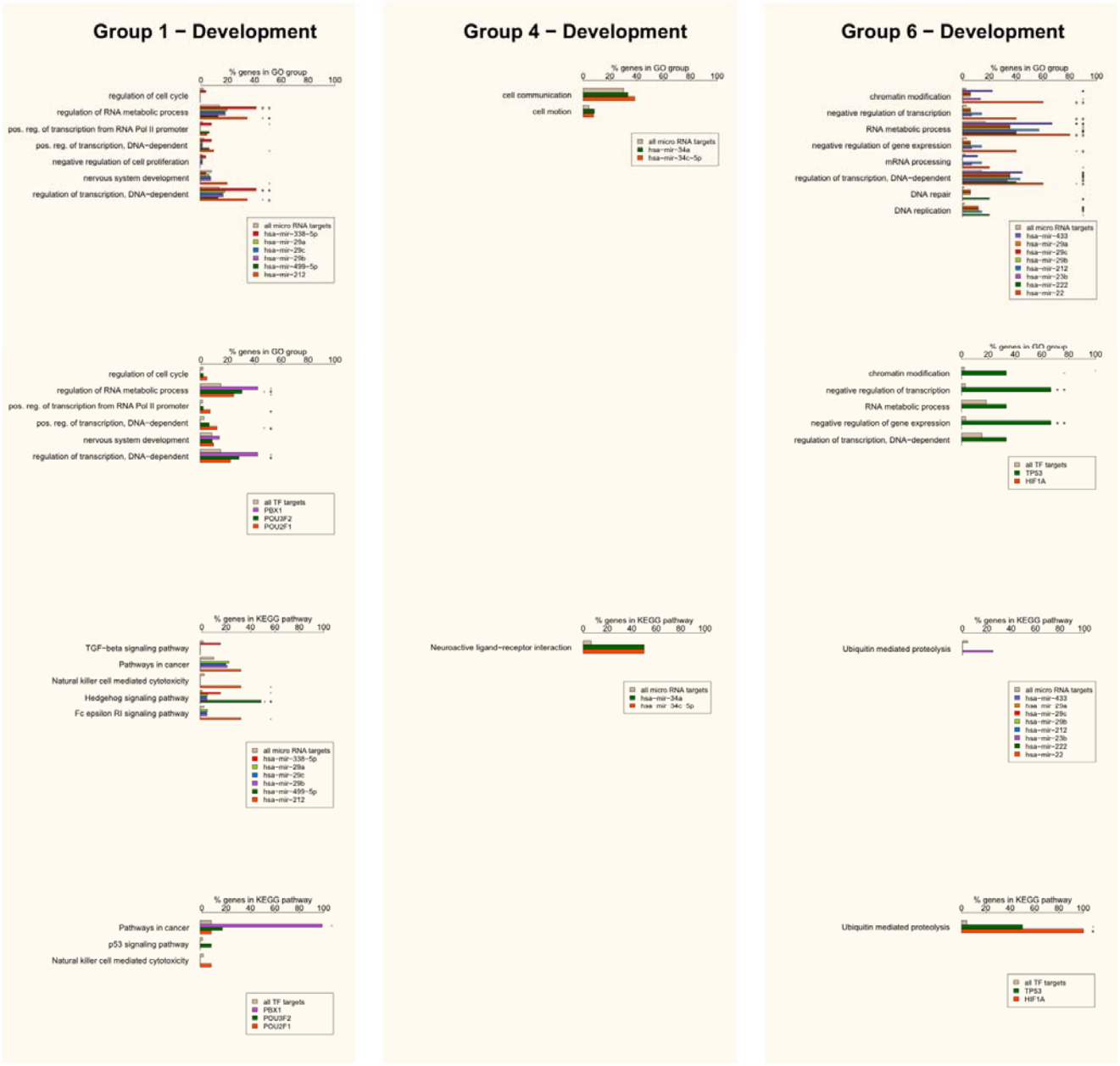


Figure S9. Transcription factor (TF) regulation of mRNA changes with age.

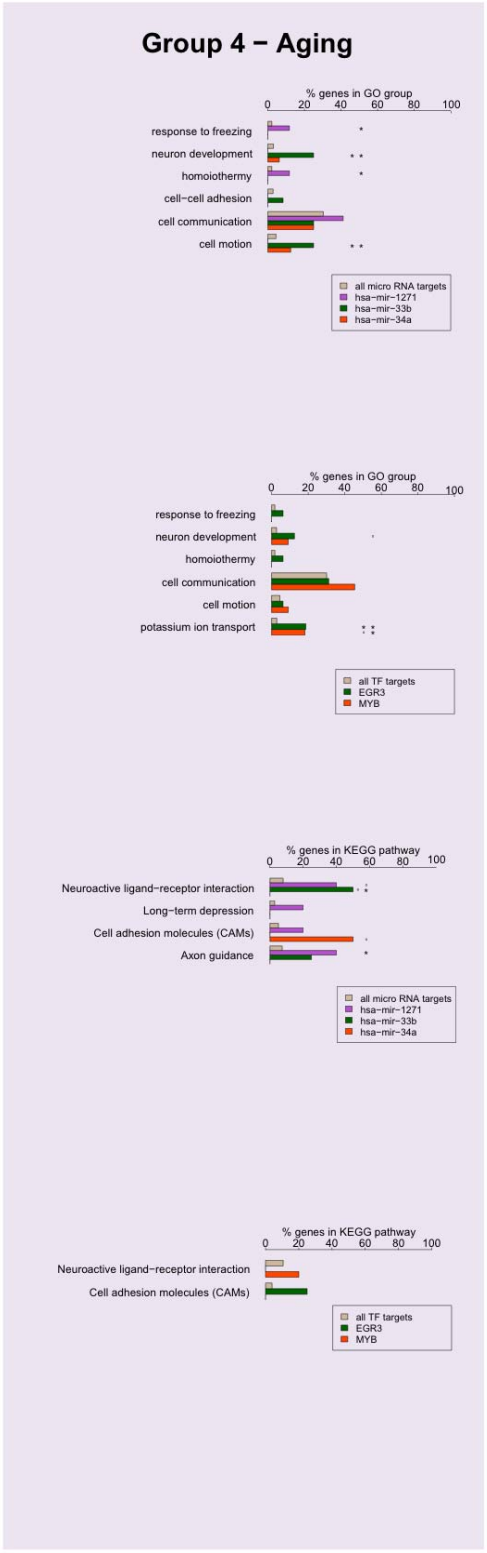
(A) Excess of positive and negative correlations between TF-target pairs in the human and rhesus macaque datasets. The two panels are drawn based on TF and their predicted target expression values using the same procedure as in Figure S8A and Figure S8B. The leftmost panels show excess of TF-target correlations in the human dataset calculated based on interpolated data. The rightmost panels show excess of TF-target correlations in the macaque, based on non-interpolated data. **(B-C)** Excess of negative or positive correlations among TF-target pairs in three selected gene groups. The y-axis indicates the proportion of TF-target pairs that show a minimum correlation. The x-axis shows the Pearson correlation cutoff. The colored bars indicate the proportion of above-cutoff correlations among TFs that show target enrichment in a gene group (at HT $p<0.05$) and their targets in that group. The grey shaded bars indicate the proportion of above-cutoff correlations among TFs *without* target enrichment in a gene group (at HT $p>=0.05$) and

their targets in that group. The TFs shown above each gene group indicate those that have significantly more negatively/positively correlated targets than in the background (at $r<-0.75$ or $r>0.75$), at binomial test $p<0.05$ (Methods). The difference between Figure S9B-C and Figure 5B is that in the latter, negative and positive correlations are combined.

A


Group 1 – Development
% genes in GO group
0 20 40 60 80 100
regulation of cell cycle
regulation of RNA metabolic process
pos. reg. of transcription from RNA Pol II promoter
pos. reg. of transcription, DNA−dependent
negative regulation of cell proliferation
nervous system development
regulation of transcription, DNA−dependent
all micro RNA targets
hsa−mir−338−5p
hsa−mir−29a
hsa−mir−29c
hsa−mir−29b
hsa−mir−499−5p
hsa−mir−212
% genes in GO group
0 20 40 60 80 100
regulation of cell cycle
regulation of RNA metabolic process
pos. reg. of transcription from RNA Pol II promoter
pos. reg. of transcription, DNA−dependent
nervous system development
regulation of transcription, DNA−dependent
all TF targets
PBX1
POU3F2
POU2F1
% genes in KEGG pathway
0 20 40 60 80 100
TGF−beta signaling pathway
Pathways in cancer
Natural killer cell mediated cytotoxicity
Hedgehog signaling pathway
Fc epsilon RI signaling pathway
all micro RNA targets
hsa−mir−338−5p
hsa−mir−29a
hsa−mir−29c
hsa−mir−29b
hsa−mir−499−5p
hsa−mir−212
% genes in KEGG pathway
0 20 40 60 80 100
Pathways in cancer
p53 signaling pathway
Natural killer cell mediated cytotoxicity
all TF targets
PBX1
POU3F2
POU2F1
Group 4 – Development
% genes in GO group
0 20 40 60 80 100
cell communication
cell motion
all micro RNA targets
hsa−mir−34a
hsa−mir−34c−5p
% genes in KEGG pathway
0 20 40 60 80 100
Neuroactive ligand−receptor interaction
all micro RNA targets
hsa−mir−34a
hsa−mir−34c−5p
Group 6 – Development
% genes in GO group
0 20 40 60 80 100
chromatin modification
negative regulation of transcription
RNA metabolic process
negative regulation of gene expression
mRNA processing
regulation of transcription, DNA−dependent
DNA repair
DNA replication
all micro RNA targets
hsa−mir−433
hsa−mir−29a
hsa−mir−29c
hsa−mir−29b
hsa−mir−212
hsa−mir−23b
hsa−mir−222
hsa−mir−22
% genes in GO group
0 20 40 60 80 100
chromatin modification
negative regulation of transcription
RNA metabolic process
negative regulation of gene expression
regulation of transcription, DNA−dependent
all TF targets
TP53
HIF1A
% genes in KEGG pathway
0 20 40 60 80 100
Ubiquitin mediated proteolysis
all micro RNA targets
hsa−mir−433
hsa−mir−29a
hsa−mir−29c
hsa−mir−29b
hsa−mir−212
hsa−mir−23b
hsa−mir−222
hsa−mir−22
% genes in KEGG pathway
0 20 40 60 80 100
Ubiquitin mediated proteolysis
all TF targets
TP53
HIF1A

B


Group 4 − Aging
% genes in GO group
0 20 40 60 80 100
response to freezing
neuron development
homoiothermy
cell−cell adhesion
cell communication
cell motion
all micro RNA targets
hsa−mir−1271
hsa−mir−33b
hsa−mir−34a
% genes in GO group
0 20 40 60 80 100
response to freezing
neuron development
homoiothermy
cell communication
cell motion
potassium ion transport
all TF targets
EGR3
MYB
% genes in KEGG pathway
0 20 40 60 80 100
Neuroactive ligand−receptor interaction
Long−term depression
Cell adhesion molecules (CAMs)
Axon guidance
all micro RNA targets
hsa−mir−1271
hsa−mir−33b
hsa−mir−34a
% genes in KEGG pathway
0 20 40 60 80 100
Neuroactive ligand−receptor interaction
Cell adhesion molecules (CAMs)
all TF targets
EGR3
MYB

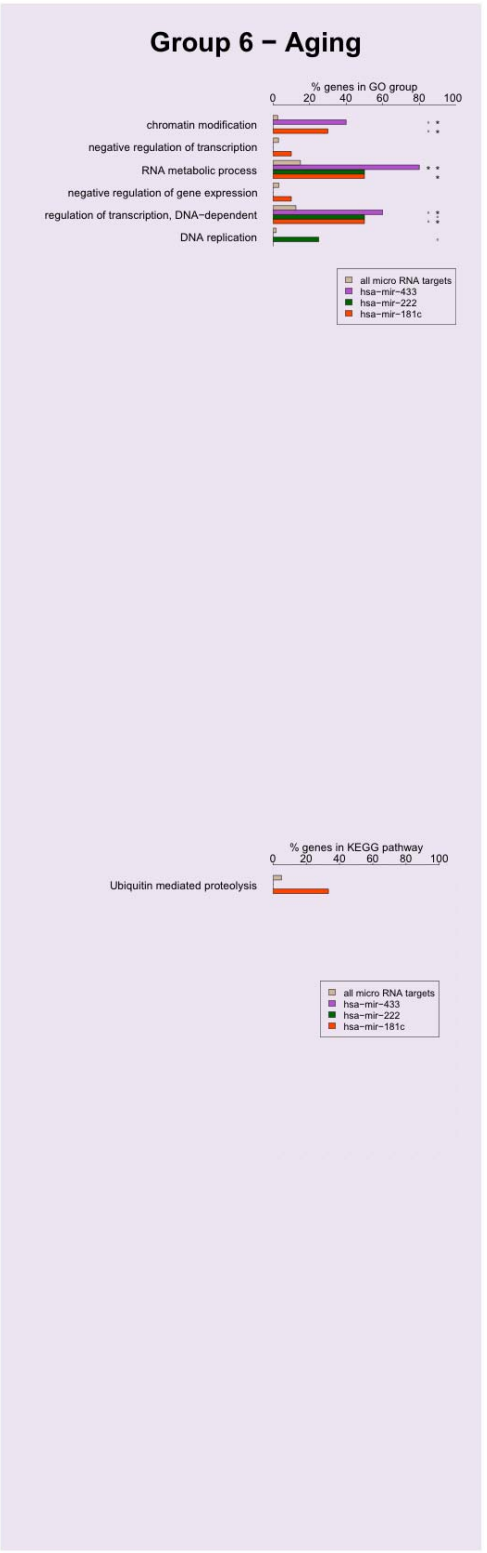

Group 6 − Aging
% genes in GO group
0 20 40 60 80 100
chromatin modification
negative regulation of transcription
RNA metabolic process
negative regulation of gene expression
regulation of transcription, DNA−dependent
DNA replication
all micro RNA targets
hsa−mir−433
hsa−mir−222
hsa−mir−181c
% genes in KEGG pathway
0 20 40 60 80 100
Ubiquitin mediated proteolysis
all micro RNA targets
hsa−mir−433
hsa−mir−222
hsa−mir−181c

Figure S10. Gene Ontology and KEGG categories enriched in miRNA and TF targets.

We tested gene sets targeted by putative miRNA regulators for enrichment in Gene Ontology biological process categories (Ashburner et al. 2000) or KEGG pathways (Kanehisa et al. 2008). GO and KEGG categories were pre-selected based on their enrichment in the respective gene groups (Table S7 and Table S8, Methods). These gene sets were compared to two different backgrounds: (a) all 4,084 age-related genes (asterisks on the right), (b) genes in the particular gene group (asterisks on the left). The significance level labels indicate *: $p<0.05$, ^{o}: $p<0.10$. The results are calculated separately for development **(A)** and aging **(B)**.

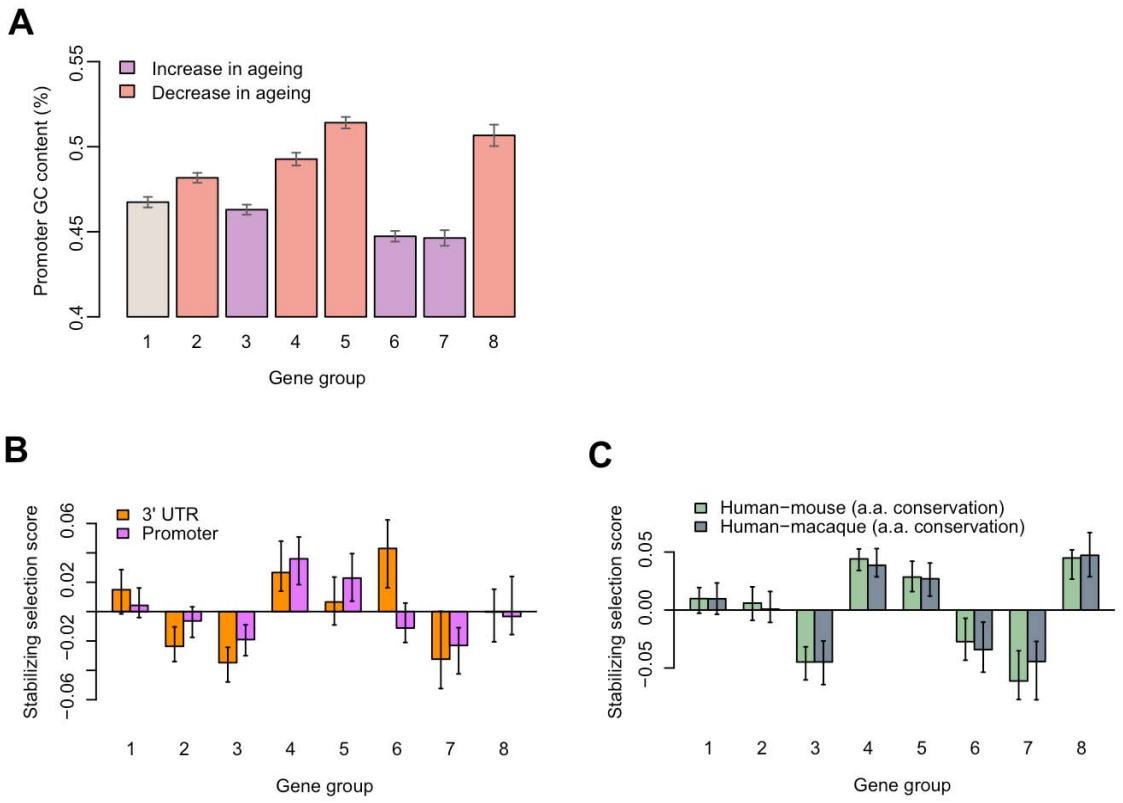


Figure S11. Promoter GC and phylogenetic conservation levels among gene groups. **(A)** The average promoter GC content among genes in the 8 co-expressed gene groups. The GC content was calculated for +/- 2000 kb of the transcription start site of each Ensembl gene. The error bars show the standard error of the mean. **(B-C)** Phylogenetic conservation among age-related genes in the 8 co-expressed mRNA groups. **(B)** Median 3' UTR and promoter conservation scores among gene groups. The conservation scores are calculated per gene as average Phastcons scores in 3' UTR or promoter regions, divided by intronic conservation scores to control for mutation rate differences (Methods). The y-axis shows the z-transformed median conservation score across gene groups. Positive values indicate above-average conservation within gene groups. In both panels, vertical bars indicate 95% confidence intervals estimated by 1,000 bootstraps across genes within groups. **(C)** Amino acid sequence conservation rates among gene groups. The y-axis shows the z-transformed median conservation score (negative dN/dS) across groups. The rates are calculated between human and mouse, and human and rhesus macaque genomes, obtained from Ensembl (Hubbard et al. 2007).

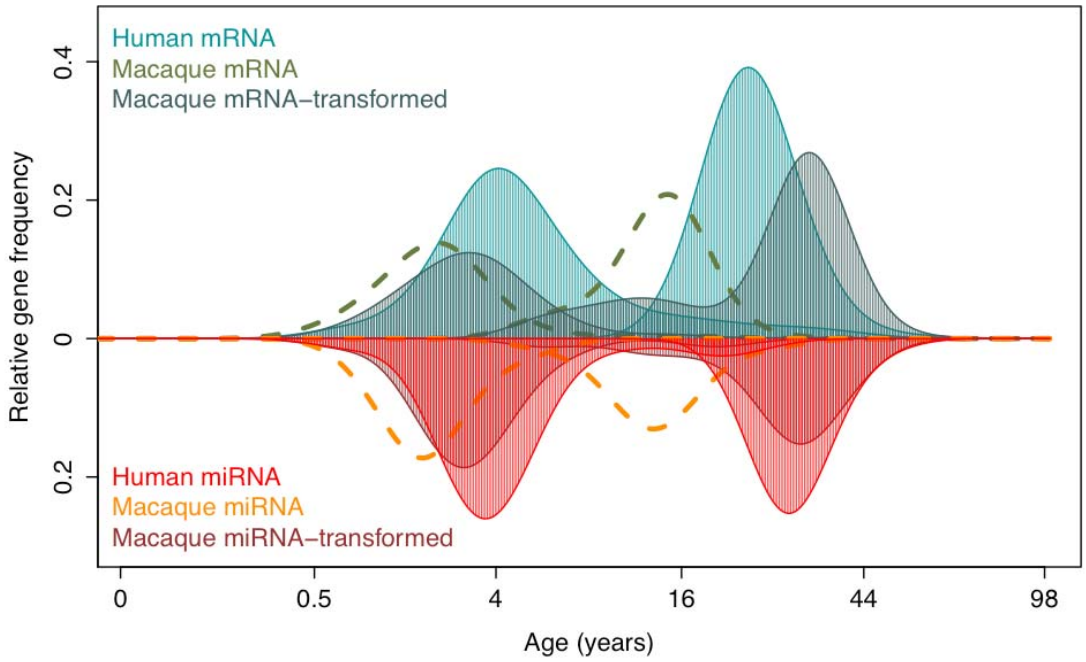


Figure S12. Match between human and macaque transition points of expression change with age.

The y-axis shows relative frequency of transition points across age-related genes, calculated using Gaussian kernel density estimation (using the "density" function in R). The x-axis shows transition points of expression change (see Figure 3), in $(age)^{1/4}$ scale. Here we estimate macaque transition points while transforming macaque ages based on a regression of life-history traits between human and rhesus macaque (Text S1). This yields the dark blue and dark red curves for macaque mRNA and miRNA, respectively. Dashed lines: transition point distributions calculated using the original macaque ages. The red and blue curves are human transition point distributions for mRNA and miRNA, respectively. Note that the age-transformation yields similar transition point distributions between the two species.

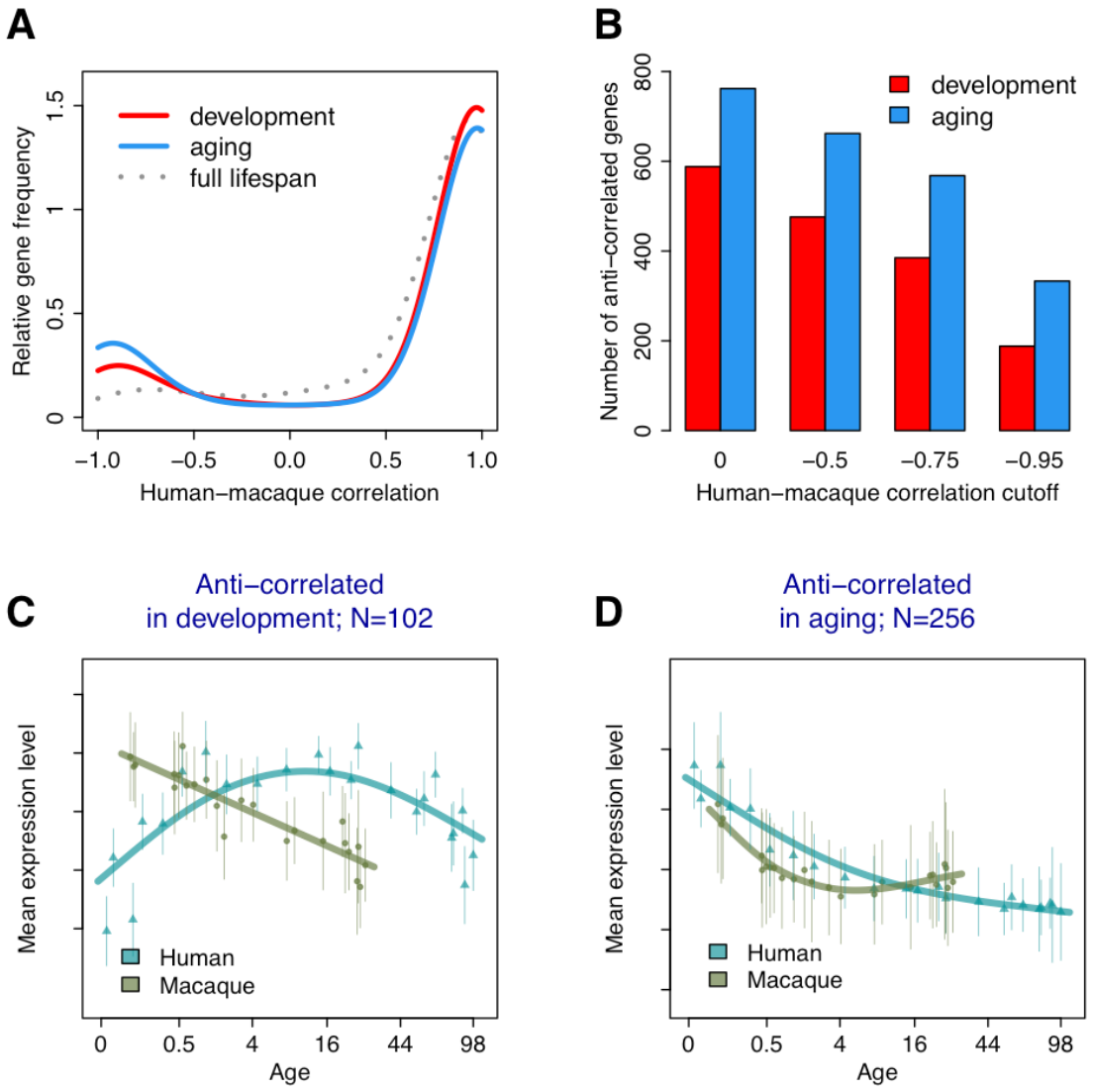


Figure S13. Human-macaque differences in development and aging.

(A) Correlation of expression changes between human and rhesus macaque during brain development (<20 years of age for human, <4 years of age in macaque; red curve), and aging (>20 and >4, respectively; blue curve). The grey dotted line indicates correlation across the full lifespan. Pearson correlations are calculated after transforming macaque ages based on a regression of life-history traits between human and rhesus macaque (Text S1), calculating cubic spline curves using (age) $^{1/4}$ scale, and interpolating points per species. The y-axis shows the relative frequency of correlation values across all age-

related human genes with macaque orthologs (n=3,233), calculated using Gaussian kernel density estimation (using the "density" function in R). **(B)** The number of anti-correlated genes at different cutoffs of Pearson correlation between human and macaque, in development and aging. This is based on the same results as in panel A. **(C)** Anti-correlated genes between human and macaque in development. Genes are chosen at Pearson correlation cutoff $r<-0.75$ and clustered into 6 groups using k-means clustering. The largest group of anti-correlated genes is shown. The y-axis shows the mean standardized expression levels of genes in a group, the x-axis shows age of individuals, in $(\text{age in days})^{1/4}$ scale. Vertical bars show the standard deviation across genes in a group. Each point represents an individual. Light blue: human, green: macaque. Note that expression changes in the later phase of life are similar. **(D)** Anti-correlated genes between human and macaque during aging. We use the same procedure to choose and plot genes as in panel **C**.

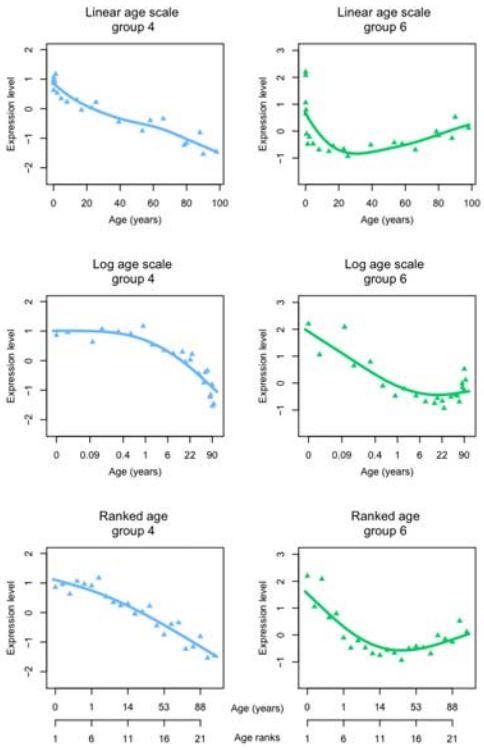


Figure S14. Modeling expression changes with alternative age scales.

Mean normalized expression levels of genes in co-expressed groups 4 and 6 in the human mRNA dataset, plotted against the age of individuals on different scales. Note the non-uniform distribution of subjects across the age scale when linear or log-transformed age scales are used.

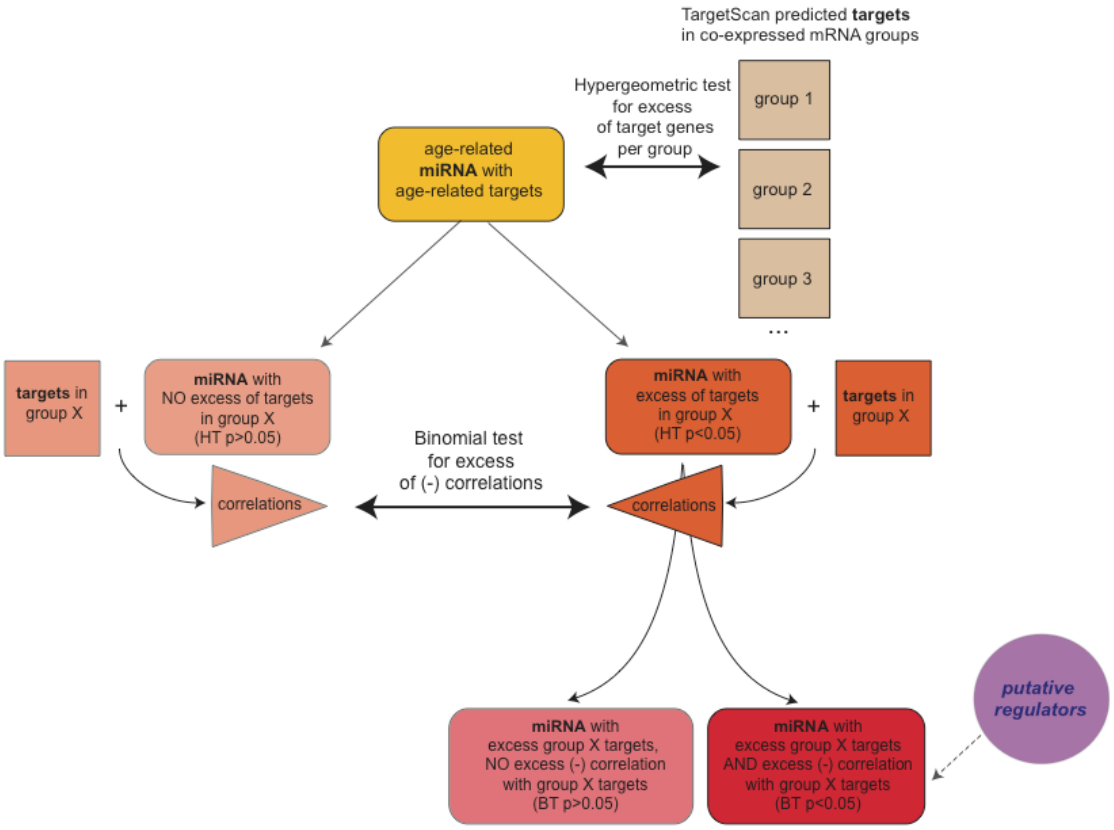


Figure S15. Scheme for identifying miRNA regulators of mRNA change.

Text S1

The R code used in the analyses is available at http://www.picb.ac.cn/Comparative/data.html.

Sample collection and RNA isolation.

Healthy human tissue was obtained from the NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland, Baltimore, MD, USA, and the Chinese Brain Bank Center, Wuhan, PR China. Rhesus macaque samples were obtained from the Suzhou Experimental Animal Center, Suzhou, PR China.

The role of the NICHD Brain and Tissue Bank is to distribute tissue and, therefore, cannot endorse the studies performed or the interpretation of results. Informed consent for use of the human tissues for research was obtained in writing from all donors or the next of kin. All subjects were defined as normal controls by forensic pathologists at the NICHD Brain and Tissue Bank. No subjects who suffered a prolonged agonal state were included.

All samples were taken from the frontal part of the superior frontal gyrus, a cortical region approximately corresponding to Brodmann area 9. For all samples, dissections contained a 2:1 grey matter to white matter volume ratio.

Hybridization to microarrays.

Total RNA was extracted from approximately 100mg of the dissected tissue sample using the standard TriZOL® protocol with no modifications, and purified with the QIAGEN® RNeasy MiniElute kit following the "RNA cleanup" protocol. RNA quality was assessed with the Agilent® 2100 Bioanalyzer system. For each sample, 2 micrograms of isolated total RNA was used as starting material for the standard Affymetrix eukaryotic target

preparation protocol (see http://www.affymetrix.com/products_services/arrays/specific/hugene_1_0_st.affx). Each sample was then hybridized to a Affymetrix® Human Gene 1.0 ST array. The samples were prepared and processed in two batches, including 12-14 human and rhesus samples each. Age distributions were similar across batches. To estimate technical variance, for each species, two individual samples were processed twice each, in both batches.

Raw and processed data was submitted to NCBI GEO and can be accessed at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18069.

Microarray data preprocessing.

We used BLAT to align all Human Gene 1.0 ST probes to the human (hg18) or rhesus macaque genomes (rhemac2) downloaded from the UCSC Genome Browser database (Karolchik et al. 2008). For calculating expression levels, we only included probes that perfectly match the respective species' genome sequence, and at a unique location.

Note that we do not use rhesus macaque genome annotation, and assume that transcript structure is conserved between the species. There will be exceptions to this, but because probes in each gene's probe set are distributed across the transcript length, we expect such instances to have only negligible effect on average expression levels used in the analysis.

We used the R Bioconductor "affy" library (Gautier et al. 2004) to extract probe intensities from the raw data files. These were then corrected for the background signal (using "antigenomic" probes of the same GC content), log-transformed, and quantile normalized. Intensity values per transcript were calculated by median polishing, following standard procedure (Gautier et al. 2004).

Criteria for probeset expression.

The following criteria were used to determine expressed transcripts: (i) In each sample, a probe with intensity >95th percentile of the antigenomic probes' intensity distribution, of the same GC content, was accepted as reliable. (ii) In each sample, a transcript was considered expressed only if it contained >7 reliable probes, and the majority of its probes were reliable. (iii) Among all samples in a dataset, a transcript was considered reliably expressed if it was expressed in ≥2 individuals per dataset. We mapped transcript IDs to Ensemble Genes using the table provided at the Affymetrix support site ("HuGene-1_0-st-v1.na26.hg18.transcript.csv"). For genes with multiple transcripts, we chose the transcript with the highest mean expression level as representative. This step avoids multiple representation of the same gene in downstream analysis.

Normalization of batches.

For both human and macaque, experimental samples were processed in two batches containing 12-14 individuals each, with similar ages. Preprocessing was also conducted separately per batch. Because the sample preparation and hybridization order can have an effect on mean expression levels or variance, we removed any such differences between batches: namely, for each gene in each batch, we z-transformed the expression levels (normalized the mean to 0 and variance to 1) and thereafter merged the datasets for each species.

miRNA isolation, sequencing, and quantification.

The procedures for microRNA experiments and pre-analyses are described in (Hu et al.

2009). Briefly, low molecular weight RNA was isolated, ligated to adapters, amplified, and sequenced following the Small RNA Sample Preparation Protocol (Illumina, USA) without modification. Technical replication was completed by independently processing the samples starting from the low molecular weight RNA isolation step.

For data preprocessing, adapter sequences were trimmed from the 3'-end of the reads as described in (Hu et al. 2009). Trimmed sequences (18-26nt long) were mapped to the human genome (hg18) using the Illumina-supplied ELAND algorithm, requiring perfect match. Per sample, ~60% of sequence reads could be mapped, whereas <1% are expected by chance, as determined by mapping scrambled reads. To annotate and quantify miRNAs, we used miRBase version 11 (Griffiths-Jones et al. 2006), only including sequences with copy number ≥2. We required all sequences to map within three nucleotides up- / down-stream of the annotated 5'-position of the mature miRNAs. For each mature miRNA, the sequence with the maximal copy number was designated as the reference sequence. The expression level of each miRNA was then calculated as a sum of the copy number of the reference sequence and the sequences mapping at the same 5'-end position as the reference sequence. Raw and processed data was submitted to NCBI GEO and can be accessed at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18069.

Novel miRNA identification.

For the miRNA precursors with one annotated miRNA, small RNA sequences mapping to the opposite arm of the precursor hairpin were also included in the analysis as novel miRNA (Hu et al. 2009). The sequence with the maximal copy number was considered a novel miRNA candidate. As a further criterion, we required the existence of at least 14 bp overlap between the annotated miRNA and the novel miRNA candidate within the

precursor hairpin. The quantification process for novel miRNA was the same as for known miRNA.

Rhesus macaque miRNA identification.

For quantifying macaque miRNAs, we used human miRNA sequences downloaded from miRBase (v.11) (n=944). First, for each human miRNA, we extracted the most likely miRNA precursor ortholog in macaque. Specifically, we mapped all annotated human miRNA precursors to the rhesus macaque genome (rhemac2) using reciprocal BLAST with the following parameters: [-F F -b 1 –e 10^{-5}]. We further required the length of hit sequence to be longer than 70% and shorter than 130% of the query sequence. Next, we extracted mature miRNAs based on the aligned precursor sequences using the Clustalw2 alignment program, with default parameters as in (http://www.ebi.ac.uk/clustalw/). Using an alternative software, Muscle (http://www.ebi.ac.uk/muscle/), yields essentially the same results (data not shown). We thus identified 776 orthologous miRNA in macaque, 306 of which are detected (total expression $\geq$100 reads) in our expression dataset. Note that highly expressed miRNA, which we use in downstream analysis, tend to be highly conserved at the sequence level (Liang and Li 2009). Consistently, among all orthologs, 76% of identified sequences share identical mature sequences between human and macaque, and among the 306 detected miRNAs, this proportion is 97% (Figure S8F). Quantification of macaque miRNA expression levels was performed as for human expression levels.

Protein sample preparation, sequencing and peptide identification.

For extraction, we used 100mg frozen prefrontal cortex samples from 12 human

individuals (Table S1). The 12 human samples were processed in two batches with 6 individuals in each, with similar age distributions in both batches. Tissue samples were minced, washed in ice-cold PBS and homogenized in ice-cold lysis buffer (8 M urea, 4% CHAPS, 65 mM DTT, 40 mM Tris, cocktail protease inhibitor, 100 mg of tissue/1 ml) using an electric homogenizer. The resulting protein solutions were sonicated on ice for a total of 3 minutes and then centrifugated at 25,000g for 1 hour at 4°C to remove DNA, RNA and other cell debris. Next, the protein supernatants were precipitated using 5× volumes of precipitation solution (ethanol: acetone: acetic acid = 50:50:0.1, volume ratio) at 4°C overnight, centrifugated and resolubilized in denaturing buffer [6 M guanidine hydrochloride, 100mM Tris, cocktail protease inhibitor, phosphatase inhibitors (1mM sodium orthovanadate and 1mM sodium fluoride), pH 8.3]. Protein concentration was determined using the Bradford assay. Next, 600μg of protein from each sample was reduced with DTT (100μg / 1μl 1M DTT), alkylated with IAA (100μg / 2μl 1M IAA), and precipitated again at 4°C overnight (as described above). After centrifugation, the resulting precipitates were resolubilized in digestion buffer (100mM ammonium bicarbonate) and incubated with Trypsin (enzyme:protein = 1:40, mass ratio) at 37°C for 20 hours, followed by ultrafiltration and lyophilization.

Each peptide sample was resolubilized in 50μl SCX loading buffer, loaded on a SCX (Strong Cation Exchange) column (Column Technology Inc., CA, USA) and eluted using a pH continuous gradient buffer (from pH 2.5-8.5), resulting in 10 fractions. Each of these fractions was then automatically loaded on one of two RP (Reversed Phase) alternative trap columns, by switching to the other RP column every 3 hours. Notably, the pH gradient-based elution yields more basic peptides than the commonly used salt-based

elution and also increases sequence coverage and concomitant confidence levels in protein identification (Zhou et al. 2007). Analysis was performed on the LTQ mass spectrometer equipped with a metal needle electrospray interface mass spectrometer (ThermoFinnigan, San Jose, CA, USA) in a data-dependent collection model (each full scan followed by ten MS/MS scans of most intense ions).

The peptides were identified by searching against an IPI human database (IPI human v3.61) using the SEQUEST program in BioWorks™ 3.2 software suite. We allowed a mass tolerance of 3.0Da and one missed cleavage site of trypsin. Cysteine carboxyamidomethlation was set as static modification and methionine oxidation was set as variable modification. All output results were filtered and integrated to proteins by an in-house software BuildSummary, using common filtering parameters: Xcorr 1.9 for charge +1; Xcorr 2.2 for charge +2; Xcorr 3.75 for charge +3 and delta CN above 0.1. By reversed database searching, at a cutoff minimum total of 20 peptides among all 12 samples, the FDR of protein identification was estimated to be 3%.

Protein data preprocessing.

Peptide identification in the two batches, containing 635,527 and 665,515 total peptide counts, were preprocessed separately and subsequently joined together. Peptide IDs were mapped to human Ensembl gene IDs using the Ensembl database (v. 55); peptides mapping to multiple genes were excluded (188 and 208 peptide IDs out of 7,551 and 6,996 in the two batches, respectively). If >1 peptide IDs were assigned to an Ensembl gene, we calculated expression levels as the mean across peptides. Peptide counts were base-two log transformed and quantile normalized for downstream analysis. We restricted downstream analysis to proteins with total peptide counts ≥20. The protein expression

level dataset is available from [http://www.picb.ac.cn/Comparative/data_methods/age_reg_2009/pr_h_cx1.txt].

Age-scales for modeling expression change.

A $\log_2$-age scale is suited for modeling exponential changes during early development (e.g. (Clancy et al. 2001; Lu et al. 2001; Shupe et al. 2006)). In contrast, linear age allows better identification of linear changes that occur during aging (e.g. (Rodwell et al. 2004; Erraji-Benchekroun et al. 2005)). To avoid bias toward development or ageing, we used age ranks of the samples to simultaneously capture developmental and aging-related changes (Figure S14). Using age ranks also ensures more uniform distribution of samples across age scale, compared to the other two scales. At the same time, it limits heteroscedasticity across ages, which pertains to the assumption of parametric regression models that errors be normally distributed and homogeneous among samples (Sokal and Rohlf 1995). We therefore use age-rank-based estimates when testing for age-effects. For comparison between two species, we use the $\log_2$-age scale, unless otherwise indicated. The choice of age scale does not qualitatively affect our results regarding development-aging reversal or regulation (data now shown).

Variance explained by age and other factors.

For calculating the average expression variance explained by age, we followed (Somel et al. 2009). Briefly, per gene, we fit the following formula:

$$Y_{ij} = \beta_{0i} + \beta_{1i} A_j + \beta_{2i} A_j^2 + \beta_{3i} A_j^3 + \varepsilon_{ij}, \quad (1)$$

where Y_{ij} is the expression level for gene i and subject j, A_j is the age of the subject j (transformed to ranks; see above and Figure S14) and ε_{ij} is the error term. The sum of

squared errors in this model are compared to the null model, $Y_{ij} = \beta_{0i} + \varepsilon_{ij}$.

For the mRNA microarray datasets, the test was performed on each batch separately. We estimated the significance of the proportion of variance explained by age by randomly permuting ages of the subjects and repeating the calculation 300 times.

Using the same scheme, the proportion of expression variance explained by sex, PMI or RIN was calculated as approximately 8%, 30% and 30%, respectively. However, these proportions were not significantly different from random (permutation test [PT] $p>0.05$). In contrast, age explains 65% and 61% of variance in both batches, while 25% and 27% are expected by chance (PT $p<0.01$). Note that neither sex, PMI nor RIN are correlated with individual age.

Comparing human and macaque. To estimate conservation of age-related expression change trends in the two species, we calculated Pearson correlations between expression profiles of orthologs. The genes were chosen to show significant expression change with age, based on the human dataset (4,084 mRNA and 115 miRNA, at FDR<0.1%). The correlations were based on interpolated points from spline curves that describe expression change with age (degrees of freedom=3), estimated for each gene in each pair of datasets and calculated for the overlapping age window in both datasets.

Clustering genes in groups.

We grouped age-related genes into 8 co-expressed groups using the human mRNA dataset of age-related genes (4,084 X 23) using k-means clustering. Before clustering, each gene was standardized to mean=0 and standard deviation=1. Because k-means is a heuristic algorithm, we repeated the procedure 10,000 times to determine the most

frequent constellation (which was 28%, while the next most frequent constellation was 10%), and used this clustering in downstream analyses. We also performed clustering with 12 groups, and repeated the downstream analyses in order to identify the extent our results are affected by the initial choice of k (Figure S3). By plotting individual genes' expression profiles, and by calculating the correlation between each group's average profile and member genes' profiles, we confirmed that the group averages are representative of their members (Figure S2).

Transition point analysis.

We used the following procedure, based on segmental linear regression, to estimate the age when early- and late-life expression change trends intersect:

(i) Fit spline curves to the expression data to describe change with age (with variable degrees of freedom, as described in (Somel et al. 2009)).

(ii) Interpolate 25 points along the age range. Interpolation is done because our sample's ages are not uniform along the age-range, and interpolating renders transition point prediction immune to such irregularities.

(iii) Divide the age-range into two sections at each subpoint within this range (excluding the most extreme ones).

(iv) For the two sections separately, construct two linear regression functions describing expression change with age (see Figure 3B).

(v) Among all subpoints, choose the one at which the two functions explain maximum variance as the putative "transition point".

(vi) On the original (not interpolated) data, test if fitting two regression models, compared to fitting a single linear regression model, is significantly better. The comparison is done

with the F-test.

(vii) Choose "transition genes" that are both age-related (using the age-test described above), to which a polynomial regression model fits at least as good as a linear regression model, and which show significant (F-test $p<0.05$) support for a transition point [as described in (vi)].

The whole procedure was performed using a log transformed age scale to identify transition points in the early developmental period (e.g. Figure 3B), as well as using the linear age scale to identify transition points in the post-developmental period. Figure 3A represents both estimates. The proportions of age-related genes/proteins chosen as "transition genes" range from 25% in the human protein dataset to 89% in the human mRNA dataset (Table S3). See below for a more information on the age range in transition point analysis.

Age ranges in transition point analysis.

Note that, in order to estimate a transition point accurately, we need sufficient information about expression change both before and after the candidate point. Therefore, it is advisable to limit a search to a range with sufficient numbers of data for earlier and older ages. Thus we did not interpolate the whole age-range in the datasets, but started from 2 months (the range of the spline curves in Figure 3B), and ended at 70 years.

In the analysis shown in Figure 3, for macaque, we used the same scale as human. On the other hand, in Figure S12, we compared macaque and human, while correcting for life-history differences between the species. For this, we first calculated the regression between time of first and last deciduous tooth eruption, first and last permanent tooth eruption, female sexual maturation, first reproduction, and maximum lifespan in human

and rhesus macaque, using a number of resources (Smith et al. 1994; Walker et al. 2006a; Walker et al. 2006b; de Magalhães and Costa 2009). This yielded the following formula: $age_{human} = 2.257 + 2.953*age_{macaque}$. We then repeated the transition point identification procedure after transforming macaque ages based on this formula.

miRNA/TF binding site estimation.

For estimating conserved miRNA target sites, we used the Conserved Site Context Score Table from TargetScan5.0 (Lewis et al. 2005). The EntrezGene IDs were mapped to Ensembl gene IDs using the Ensembl database. For estimating conserved TF binding sites (TFBS), we used the Ensembl database for gene annotation, and extracted 2,000 bp +/- around the transcription start site (TSS) for each gene. We ran the Match™ algorithm, which uses the TRANSFAC® database (Kel et al. 2003), on the extracted proximal promoter regions, using the input files matrixTFP112.lib and vertebrate_non_redundant_minFP.prf. We then calculated average Phastcons (conservation) scores, using the UCSC Genome Browser 17-way vertebrate Conserved Element Table (Siepel et al. 2005), on each predicted TFBS site. For each candidate TFBS, we required that ≥80% of nucleotides have Phastcons score and an average score ≥0.6.

TFBS names were mapped to transcription factors using two paths: (a) we used HGNC symbols parsed from the Transfac® Matrix Table (v. 11.2). These were mapped to Ensembl genes using the HGNC database (www.genenames.org/data/gdlw_index.html). (b) We used the UCSC Genome Browser HMR Conserved Transcription Factors Table, which contains Transfac TFBS IDs and corresponding SwissProt IDs; the latter were mapped to Ensembl genes using the Ensembl database. The genes lists from the two

paths were combined, which resulted in 426 TFBS mapped to ≥1 Ensembl gene that were also annotated with "transcription factor activity" (GO:0003700) in Ensembl.

Regulator miRNA/TF identification – detailed description.

In order to restrict the false discovery rate of regulator-target identification, we limited the target predictions to genes showing age-related change at F-test $p=4*10^{-6}$ (the Bonferroni correction cutoff). For miRNA/TF, we required evidence for age-related change at $p<1*10^{-3}$ (a more relaxed cutoff was chosen due to the smaller number of miRNA/TF, compared to all expressed genes).

For the identification of miRNA regulators of brain development/aging, we required two conditions: **(1)** Enrichment of targets in one of the 8 co-expressed age-related gene groups. For this, we tested if a miRNA's targets are enriched among genes in a gene group using the HT, compared to other miRNAs and all other gene groups (including only genes with ≥1 target site). To avoid the influence of miRNA gene families, we restricted this comparison to a single test per unique miRNA seed. These results were compared to results obtained using randomized gene groups generated by 1,000 permutations of genes among 8 co-expressed clusters. If a miRNA's targets were enriched in a cluster at HT $p<0.05$, we considered that miRNA "specific" to that gene group. All other miRNA-targets pairs (at $p>0.05$) were considered "non-specific." **(2)** Negative correlation with targets in a specific gene group. For this, we calculated correlations between the miRNA and its targets. The correlations were separately calculated for development and aging periods (with 20 and 5 years as borderline for human and macaque, respectively, representing approximate times of age at first reproduction (Walker et al. 2006a; Walker et al. 2006b)). For both periods we use data from similar

numbers of individuals. Using these correlations, for each miRNA, for each gene group, we determined the proportion of negatively correlated miRNA-target pairs in that gene group (at correlation cutoff Pearson $r<-0.75$, $p<0.05$). Next, for each miRNA that showed excess of targets in a particular gene group (as described above), we compared the proportion of its negatively correlated targets in that gene group, with the mean proportion among all miRNAs that did not show enrichment in that gene group (which represents the random background). The proportions were compared using binomial tests (BT).

A putative regulatory miRNA was defined as one that had both a significant excess of targets in a gene group (at one-sided HT $p<0.05$) and a significant excess of negative correlation with its targets in that group (at one-sided BT $p<0.05$) (Figure S15).

Testing predicted regulators – age-independent effects.

We calculated gene expression profiles using miRNA and mRNA expression levels predicted from spline models of expression change with age (degrees of freedom=3), instead of the original expression levels. The models were calculated using log and linear age scales for estimating developmental and aging-related effects, respectively. If the miRNA-mRNA covariance is due to a factor independent of age, we expect to find no correlation in this case. Instead, the excess of negative correlations among specific miRNA-mRNA pairs became even more prominent (Figure S8A), which indicates that the correlations mainly reflect age-related changes.

Testing predicted regulators – conservation in rhesus macaque.

To identify regulators in rhesus macaque conserved between macaques and humans, we

pre-selected miRNA showing the same direction of expression change with age (in either the development or aging periods) as in human (67 and 66 miRNAs, respectively, out of 98 age-related miRNA with macaque orthologs). We then separated these miRNA-gene group pairs into "enriched" and "non-enriched" classes, and compared the correlations between the two groups, as described above. Note that regulator identification may be less precise in the rhesus macaque analysis, as it is based on target enrichment in human clusters. Furthermore, the rhesus macaque data, presumably due to its shorter age-range, shows less age-related change than human (both in development and aging, and both in miRNA and mRNA). There were 10 regulator-target gene group pairs identified for development, and 5 identified for aging (Figure 5A). In addition, we checked if putative regulator miRNA identified in human show tendency for excess negative correlations (mean $r<0$) with their targets in macaque (Table S5).

Testing predicted regulators – coordinated divergence.

Here, we tested whether miRNA expression differences between humans and macaques would be reflected in expression of their putative target genes. For instance, if a miRNA expression change occurs at a later age in humans compared to macaques, would the target genes' expression profiles be similarly shifted (Figure S8C)? We first interpolated the human and macaque expression-age curves at 25 points along each species' lifespans, for both the miRNA and mRNA datasets. Next, for each orthologous miRNA/mRNA, we calculated human-macaque expression level divergence, and standardized these to mean=0 and standard deviation=1, per gene. We then calculated the correlations between human-macaque miRNA divergence and target mRNA divergence across the 25 interpolated points. Divergence correlations at $r<-0.95$ ($p<1*10^{-6}$) were considered

"coordinated" (a stringent cutoff was chosen, as we are using interpolated data here). For putative regulatory miRNAs, which are enriched in a gene group and which are negatively correlated with their targets in human (see above), we expected to observe coordinated divergence between human-macaque. In contrast, we did not expect coordinated divergence for background miRNA-target pairs. Indeed, the proportion of coordinated divergence among putative regulator miRNA-target pairs is higher than the same proportion among miRNA not enriched for target sites in the same gene group (Figure S8D). The significance of the difference was calculated using the hypergeometric test (HT).

Testing predicted regulators – target site mutation test.

In another approach, we took advantage of the fact miRNA-based regulation requires perfect base-pairing between a 6 to 7 nucleotide stretch in miRNA 5' end, the "seed" region, and gene's 3' UTR. Due to sequence differences between the human and the rhesus macaque genomes, some of the miRNA binding sites predicted in humans are disrupted in macaques. If our predictions are correct, loss of the target sites should disrupt miRNA-target relationship, which should be reflected in lack of negative correlation between miRNA expression and expression of its target genes, or dysregulation, in rhesus macaques. We calculated the number of mutations in the macaque orthologs of human miRNA target sites. To define human miRNA target sites, we used the Conserved Site Context Score Table from TargetScan5.0 (Lewis et al. 2005). We then extracted the corresponding orthologous seed sequences in rhesus macaque from the TargetScan5.0 23-way UTR sequence alignment table. Next, we divided each seed sequence pair between human and macaque into 3 categories based on sequence similarity: (a) containing at

least one mutation (among 7,331 miRNA-target pairs, 298 (4.1%) contained at least one mutation), (b) exactly the same, (c) uncertain, if a human microRNA target site is missing in macaque. Using this data, we calculated the proportion of miRNA-target pairs that show negative correlation in human at a certain correlation cutoff, but show weaker negative correlation in macaque, which we refer to as dysregulated miRNA-target pairs. Next, using the HT, we compared the proportion of dysregulated cases among miRNA-target pairs with a mutation in macaque, with the proportion of dysregulated cases among miRNA-target pairs devoid of mutation (Figure S8E). In this test we only used miRNAs that show significant enrichment in the gene groups (see section "Regulator miRNA/TF identification"). The proportion of dysregulated pairs was calculated at different cutoffs. Disrupted target sites was found to lead to decreased correlation among specific miRNA-target pairs in macaque: the effect is significant for gene expression changes taking place in development (HT $p<0.01$), although appears only as a tendency for the regulators identified for aging (Figure S8E, note the small numbers of mutated miRNA binding sites). Hence, a loss of conserved binding sites leads to a loss of regulation, as predicted by the consensus model of miRNA function.

Testing predicted regulators – comparison with experimentally verified targets.

We compared miRNA-target gene pairs estimated in our study with four experimentally verified miRNA target gene sets. We used two databases, Tarbase (http://diana.cslab.ece.ntua.gr/tarbase/tarbase_download.php) (Papadopoulos et al. 2009) and Mirwalk (http://www.ma.uni-heidelberg.de/apps/zmf/mirwalk/contact.html, unpublished); we also used a recently published study by Khan et al. (Khan et al. 2009), which collected results from multiple experiments, and results from Baek et al. (Baek et

al. 2008) on mRNA regulation by overexpression of miR-181. All gene sets were converted to human Ensembl gene IDs using Ensembl Biomart. Table S6 lists miRNA-target gene pairs predicted in our study (based on the target site enrichment and correlation tests, shown in Table S5) that overlap with each of these four datasets. The significance of each overlap was estimated by a randomization test. For this, we used all 1,771 potential age-related target genes (based on TargetScan (Lewis et al. 2005)). Among these genes, we randomly chose and assigned pseudotargets to each miRNA; the same number as predicted targets. We then tested the overlaps between the random sets of targets, with targets in each database. This was repeated 1,000 times, per database. The *p*-value was calculated as the number of times we find an as large or larger overlap between pseudotarget sets and the database, as originally observed. The random expectation was calculated as the median number of overlapping target-gene pairs in the permutations. Note these datasets partially overlap; also, most experiments involve over-expression of a miRNA in cell lines (e.g. HeLa cells) and thus represent a quite distinct environment compared to the brain. Therefore, finding only limited overlap with these experiments is not unexpected.

FDR of regulator prediction.

We estimated the false discovery rate (FDR) in the binding site enrichment test and the regulator-target correlation test using permutation. In the binding site enrichment test, we find a total of 90 miRNAs showing enrichment in gene groups (note that one miRNA may target multiple groups, and the 90 consists of 80 unique miRNAs). Eighty-three (83) of these are enriched in 1, 4, and 6. For FDR estimation, we randomly assigned age-related genes to gene groups, and calculated the number of regulators with enriched

binding sites in these random groups, 1,000 times. Using the median number of miRNAs found across the permutations, we estimated an FDR of 24.6%. In the binomial test for higher negative correlations between miRNAs and their targets in a gene group, we find 22 miRNA-target pairs (16 in development, 6 in aging; only using age-related miRNAs). Per gene group, we then randomly chose age-related miRNAs without enriched binding sites in that group, the same number as age-related miRNAs with enriched binding sites. Next we applied the binomial test for higher negative correlations. This was repeated 1,000 times and yielded a FDR estimate of 52.3%. Note that these could be overestimates, as any non-identified regulators in our dataset should have an inflating effect on these figures.

Functional analysis.

We used the func_hyper (Prüfer et al. 2007) program to test for enrichment of gene groups in Gene Ontology (GO) categories (Ashburner et al. 2000). In addition to using the HT, func_hyper runs a permutation test to determine if the number of significant HT results is higher than expected by chance, given multiple testing. It thus calculates a "global enrichment p-value". We used the same strategy to test for enrichment among KEGG pathways (Kanehisa et al. 2008), using in-house code. The Ensembl gene-GO data were downloaded from the Ensembl database and Ensembl gene-KEGG data from the KEGG database.

For identifying cell type-specific expression, we used expression levels measured from purified mouse neurons, astrocytes and oligodendrocytes (Cahoy et al. 2008). From this dataset, for each mouse-human ortholog, we calculated the effect size of each size cell type, and assigned genes with effect size >2 as specific to that cell type (this cutoff

ensured no overlap between cell type specific gene groups). This procedure yielded 1116, 965, and 991 human genes specific to neurons, astrocytes and oligodendrocytes, respectively.

For putative targets of miRNAs/TFs identified in this study, we conducted secondary tests for functional enrichment. Here we only used the GO/KEGG groups that were previously identified as enriched in the respective gene group. This is because the gene numbers are too small to pass correction for multiple testing among all GO/KEGG categories. We used two backgrounds in these tests: (1) all age-related genes targeted by an age-related regulator, (2) all genes in that gene group targeted by an age-related regulator. Test type (1) is biased, because we pre-select categories found to be significantly enriched in this cluster compared to other age-related genes, and the targets belong to the same cluster. For test type (2), we expect no such bias.

Evolutionary conservation analysis.

We used the PhastCons 18-way Placental Mammal Conservation Track (a subset of the 28-way Placental Track) from the UCSC Genome Browser to calculate non-coding sequence conservation measures (Siepel et al. 2005), and the Ensembl database for gene coordinates. For each human gene, we computed mean sequence conservation for (a) 3' UTR, and (b) the proximal promoter, defined as 2,000 bp +/- around the transcription start site (TSS) (Xie et al. 2005). For genes with multiple transcripts, we chose the one with the largest number of exons. To estimate the basal mutation rate per gene, we used intronic conservation (excluding first introns, excluding 100 bp around each splice site, and considering only 2,500 bp at each end inside an intron, as used by (Haygood et al. 2007)). For estimating protein conservation, we used negative dN/dS ratios calculated for

mouse-human (hsa-mmus) and macaque-human (hsa-rhemac), downloaded from the Ensembl database (v. 55) (Hubbard et al. 2007). Notably, sequence conservation on protein-coding regions is the strongest among genes in co-expression groups 4, 5, and 8, which are enriched in neuron-specific genes (Figure 6F and Figure S11). Sequence conservation on regulatory regions, however, is the strongest amongst genes in groups 1, 4, 5, and 6 (Figure S11).

Stabilizing Selection Score.

This score is calculated as the Pearson correlation coefficient between the standardized expression levels of an individual, and sequence conservation levels (corrected for variation in mutation rate) among the 4,084 age-related genes in the human dataset. Note that all four types of conservation values are positively correlated with mean expression levels (Spearman rank correlation $p<0.05$). However, the Stabilizing Selection Score (SSS) is calculated between expression levels standardized across the 23 human individuals per gene. We also tested whether change in SSS is driven by particular gene groups under positive selection. We identified 507 genes potentially under positive selection, involved in immune and stress response, involved in reproduction, or with the ratio of non-synonymous to synonymous substitution rates (dN/dS)>1 (between human and macaque, or human and mouse), using the Ensembl and Gene Ontology databases (6,955 such genes in total). Excluding these genes does not affect the decrease in SSS with age (Figure 7B). Similarly, restricting the analysis to 935 age-related genes with enriched in neuronal expression profiles (i.e. with higher expression in neurons compared to glia, based on (Cahoy et al. 2008); 2,500 such genes in total), yields the same result (Figure 7C). Finally, we analyzed three potential confounding factors across age-related

genes: number of protein-protein interaction partners (based on the Human Protein Reference Database (Keshava Prasad et al. 2009)) and the number of tissues or cell types a gene is expressed in (i.e. expression breath; based on the GNF dataset which was downloaded from the Ensembl database, v.54). All three measures were positively correlated with conservation scores across genes, as expected (Duret and Mouchiroud 2000) (data not shown). However, none of these three measures showed decreasing correlation with expression levels with age (Figure 7D), in contrast to decreasing SSS.

References

Ashburner, M, Ball, CA, Blake, JA, Botstein, D, Butler, H, Cherry, JM, Davis, AP, Dolinski, K, Dwight, SS, Eppig, JT et al. 2000. Gene Ontology: Tool for the unification of biology. *Nat Genet* **25:** 25-29.

Baek, D, Villén, J, Shin, C, Camargo, FD, Gygi, SP, Bartel, DP. 2008. The impact of microRNAs on protein output. *Nature* **455:** 64-71.

Cahoy, JD, Emery, B, Kaushal, A, Foo, LC, Zamanian, JL, Christopherson, KS, Xing, Y, Lubischer, JL, Krieg, PA, Krupenko, SA et al. 2008. A transcriptome database for astrocytes, neurons, and oligodendrocytes: A new resource for understanding brain development and function. *J Neurosci* **28:** 264-278.

Clancy, B, Darlington, RB, Finlay, BL. 2001. Translating developmental time across mammalian species. *Neuroscience* **105:** 7-17.

de Magalhães, J and Costa, J. 2009. A database of vertebrate longevity records and their relation to other life-history traits. *J Evol Biol* **22:** 1770-1774.

Dogini, DB, Ribeiro, PA, Rocha, C, Pereira, TC, Lopes-Cendes, I. 2008. MicroRNA expression profile in murine central nervous system development. *J Mol Neurosci* **35:** 331-337.

Duret, L and Mouchiroud, D. 2000. Determinants of substitution rates in mammalian genes: Expression pattern affects selection intensity but not mutation rate. *Mol Biol Evol* **17:** 68-70.

Erraji-Benchekroun, L, Underwood, MD, Arango, V, Galfalvy, H, Pavlidis, P, Smyrniotopoulos, P, Mann, JJ, Sibille, E. 2005. Molecular aging in human prefrontal cortex is selective and continuous throughout adult life. *Biol Psychiatry* **57:** 549.

Gautier, L, Cope, L, Bolstad, BM, Irizarry, RA. 2004. affy--Analysis of Affymetrix genechip data at the probe level. *Bioinformatics* **20:** 307-315.

Griffiths-Jones, S, Grocock, RJ, van Dongen, S, Bateman, A, Enright, AJ. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucl Acids Res* **34:** D140-144.

Haygood, R, Fedrigo, O, Hanson, B, Yokoyama, K, Wray, GA. 2007. Promoter regions of many neural- and nutrition-related genes have experienced positive selection during human evolution. *Nat Genet* **39:** 1140-1144.

Hu, H, Yan, Z, Xu, Y, Hu, H, Menzel, C, Zhou, YH, Chen, W, Khaitovich, P. 2009. Sequence features associated with microRNA strand selection in humans and flies. *BMC Genomics* **10:** 413.

Hubbard, TJP, Aken, BL, Beal, K, Ballester, B, Caccamo, M, Chen, Y, Clarke, L, Coates, G, Cunningham, F, Cutts, T et al. 2007. Ensembl 2007. *Nucl Acids Res* **35:** D610-617.

Kanehisa, M, Araki, M, Goto, S, Hattori, M, Hirakawa, M, Itoh, M, Katayama, T, Kawashima, S, Okuda, S, Tokimatsu, T et al. 2008. KEGG for linking genomes to life and the environment. *Nucl Acids Res* **36:** D480-484.

Karolchik, D, Kuhn, RM, Baertsch, R, Barber, GP, Clawson, H, Diekhans, M, Giardine, B, Harte, RA, Hinrichs, AS, Hsu, F et al. 2008. The UCSC genome browser database: 2008 update. *Nucleic Acids Res* **36:** D773-779.

Kel, AE, Gössling, E, Reuter, I, Cheremushkin, E, Kel-Margoulis, OV, Wingender, E. 2003. MATCH: A tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res* **31:** 3576-3579.

Keshava Prasad, TS, Goel, R, Kandasamy, K, Keerthikumar, S, Kumar, S, Mathivanan, S, Telikicherla, D, Raju, R, Shafreen, B, Venugopal, A et al. 2009. Human protein reference database--2009 update. *Nucl Acids Res* **37:** D767-772.

Khan, AA, Betel, D, Miller, ML, Sander, C, Leslie, CS, Marks, DS. 2009. Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat Biotech* **27:** 549-555.

Lewis, BP, Burge, CB, Bartel, DP. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120:** 15-20.

Liang, H and Li, W-H. 2009. Lowly expressed human MicroRNA genes evolve rapidly. *Mol Biol Evol* **26:** 1195-1198.

Lu, L, Airey, DC, Williams, RW. 2001. Complex trait analysis of the hippocampus: Mapping and biometric analysis of two novel gene loci with specific effects on hippocampal structure in mice. *J Neurosci* **21:** 3503-3514.

Lu, T, Pan, Y, Kao, S, Li, C, Kohane, I, Chan, J, Yankner, BA. 2004. Gene regulation and DNA damage in the ageing human brain. *Nature* **429:** 883.

Papadopoulos, GL, Reczko, M, Simossis, VA, Sethupathy, P, Hatzigeorgiou, AG. 2009. The database of experimentally supported targets: A functional update of tarbase. *Nucl Acids Res* **37:** D155-158.

Prüfer, K, Muetzel, B, Do, H, Weiss, G, Khaitovich, PR, Pääbo, S, Lachmann, M, Enard, W. 2007. Func: A package for detecting significant associations between gene sets and ontological annotations *BMC Bioinformatics* **8:** 41.

Rodwell, GEJ, Sonu, R, Zahn, JM, Lund, J, Wilhelmy, J, Wang, L, Xiao, W, Mindrinos, M, Crane, E, Segal, E et al. 2004. A transcriptional profile of aging in the human kidney. *PLoS Biol* **2:** e427.

Shupe, JM, Kristan, DM, Austad, SN, Stenkamp, DL. 2006. The eye of the laboratory mouse remains anatomically adapted for natural conditions. *Brain Behav Evol* **67:** 39-52.

Siepel, A, Bejerano, G, Pedersen, JS, Hinrichs, AS, Hou, M, Rosenbloom, K, Clawson, H, Spieth, J, Hillier, LW, Richards, S et al. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15:** 1034-1050.

Smirnova, L, Gräfe, A, Seiler, A, Schumacher, S, Nitsch, R, Wulczyn, FG. 2005. Regulation of miRNA expression during neural cell specification. *Eur J Neurosci* **21:** 1469-1477.

Smith, BH, Crummett, TL, Brandt, KL. 1994. Ages of eruption of primate teeth: A compendium for aging individuals and comparing life histories. *Am J Phys Anthropol* **37:** 177-231.

Sokal, RR and Rohlf, FJ. 1995. *Biometry*. Freeman, New York.

Somel, M, Franz, H, Yan, Z, Lorenc, A, Guo, S, Giger, T, Kelso, J, Nickel, B, Dannemann, M, Bahn, S et al. 2009. Transcriptional neoteny in the human brain. *Proc Natl Acad Sci USA*.

Walker, R, Burger, O, Wagner, J, Von Rueden, CR. 2006a. Evolution of brain size and juvenile periods in primates. *J Hum Evol* **51:** 480-489.

Walker, R, Gurven, M, Hill, K, Migliano, A, Chagnon, N, De Souza, R, Djurovic, G, Hames, R, Hurtado, A, Kaplan, H et al. 2006b. Growth rates and life histories in twenty-two small-scale societies. *Am J Hum Biol* **18:** 295-311.

Xie, X, Lu, J, Kulbokas, EJ, Golub, TR, Mootha, V, Lindblad-Toh, K, Lander, ES, Kellis, M. 2005. Systematic discovery of regulatory motifs in human promoters and 3'UTRs by comparison of several mammals. *Nature* **434:** 338.

Zhou, H, Dai, J, Sheng, Q-H, Li, R-X, Shieh, C-H, Guttman, A, Zeng, R. 2007. A fully automated 2-D LC-MS method utilizing online continuous pH and RP gradients for global proteome analysis. *Electrophoresis* **28:** 4311-4319.