Supporting Information

Transcriptional neoteny in the human brain

Mehmet Somel*, Henriette Franz, Zheng Yan, Anna Lorenc, Song Guo, Thomas Giger, Janet Kelso, Birgit Nickel, Michael Dannemann, Sabine Bahn, Maree J. Webster, Cynthia S. Weickert, Michael Lachmann, Svante Pääbo and Philipp Khaitovich*

*To whom correspondence should be addressed: somel@eva.mpg.de or khaitovich@eva.mpg.de

Contents

Supporting Figures

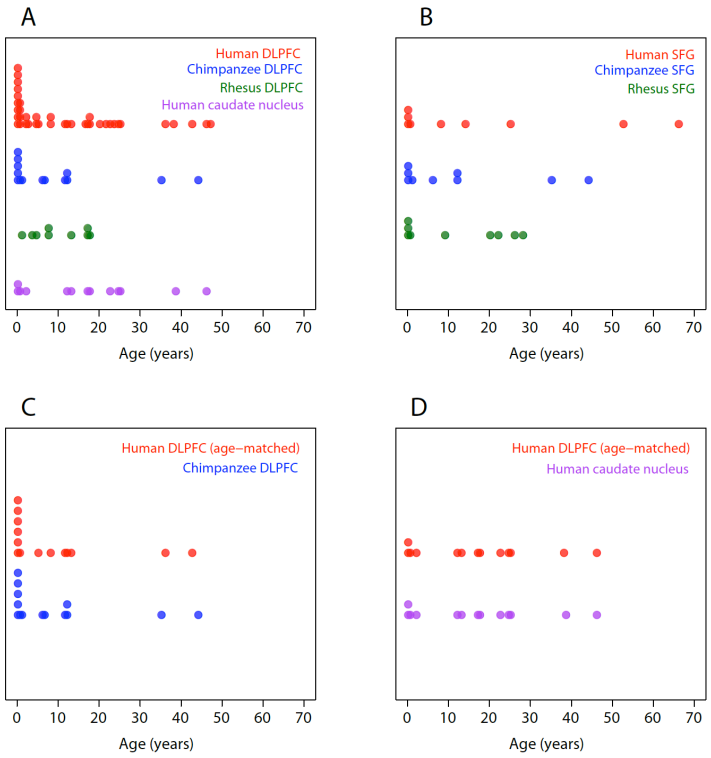

Fig. S1. Age-distributions of the subjects used in the study

Each dot represents an individual sample. The x-axis shows individuals' ages in years after birth. The colours represent: red – human prefrontal cortex [dorsolateral prefrontal cortex (DLPFC) or superior frontal gyrus (SFG)], blue – chimpanzee prefrontal cortex, green – rhesus macaque prefrontal cortex, purple – human caudate nucleus. *(A)* The distributions of ages of all primates used with the HG-U133Plus2.0 microarrays. *(B)* Age distributions of all primates used with the HuGene-1.0 ST microarrays. *(C-D)* Age-distributions of age-matched sample sets, which were used in the analysis to ensure that sample size and age distribution differences between groups does not bias the results. *(C)* Human DLPFC samples numbered 1, 2, 3, 4, 5, 8, 13, 19, 20, 22, 23, 24, 35, 37 in Table S1 were chosen to match the chimpanzee DLPFC samples in age. *(D)* Human DLPFC samples numbered 7, 9, 13, 15, 23, 24, 26, 27, 31, 33, 34, 36, 38 in Table S1 were chosen to match the human caudate nucleus samples in age.

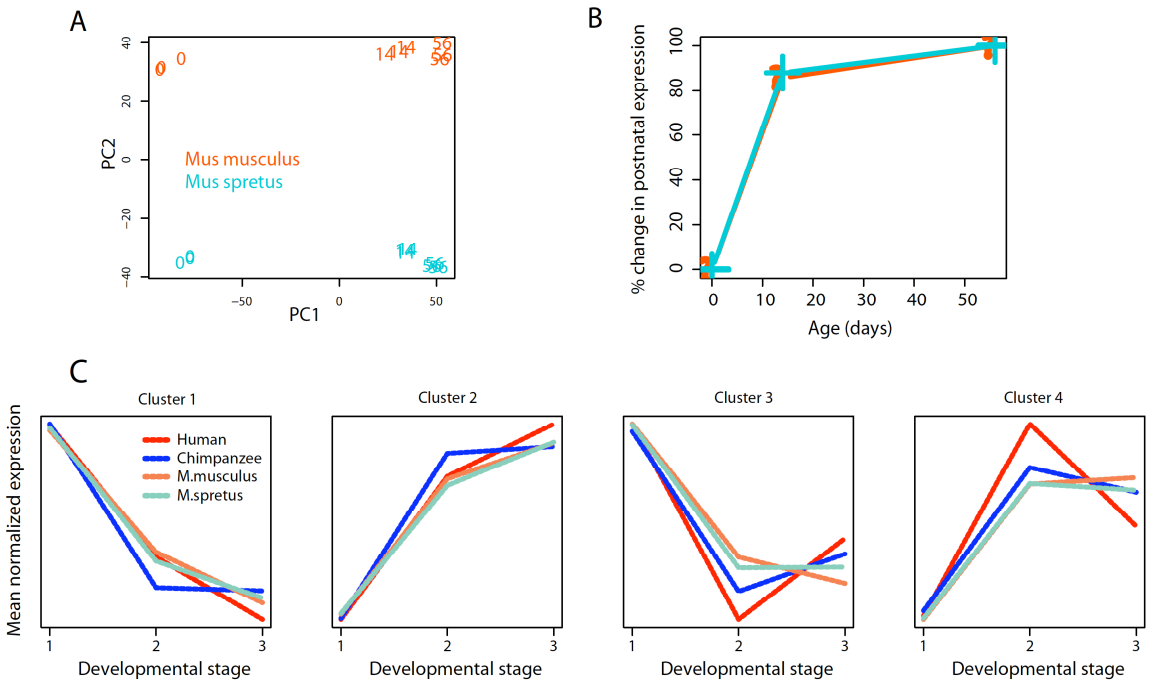

Fig. S2. Conservation of developmental patterns between mice and primates.

(A) The first two principle components of the total expression variation analysis among *M.musculus and M.spretus* individuals, based on 8,362 expressed genes. The numbers represent each individuals' birth age in days. The first and second components explain 45% and 15% of the total variation, and are correlated with age (r=0.94, $p<10^{-3}$) and species (r=0.87, $p<10^{-3}$), respectively. The colours represent different species: orange – *M.musculus* and light blue – *M.spretus*. *(B)* Percent global expression change relative to the newborns. 100% change was designated as the difference between the youngest and oldest individuals in terms of the summary measure of global expression (see SI section "The influence of age, species identity and sex on expression"). Orange lines represent *M.musculus* and light blue *M.spretus*. *(C)* Conservation of expression changes during postnatal brain development among mammals. The y-axis shows mean normalized expression levels among 2,599 human-mouse orthologs in four different species at three developmental stages. The genes are grouped into 4 k-means clusters, using expression levels of the 3 human developmental groups. Red – humans, blue – chimpanzees, orange – *Mus musculus*, light blue – *Mus spretus*. For primates, these stages are: newborns (0-1 years of age), adolescents (10-14 years) and young adults (25-50 years). For mice, these stages are: newborns (0 days), pups (14 days) and young adults (56 days).

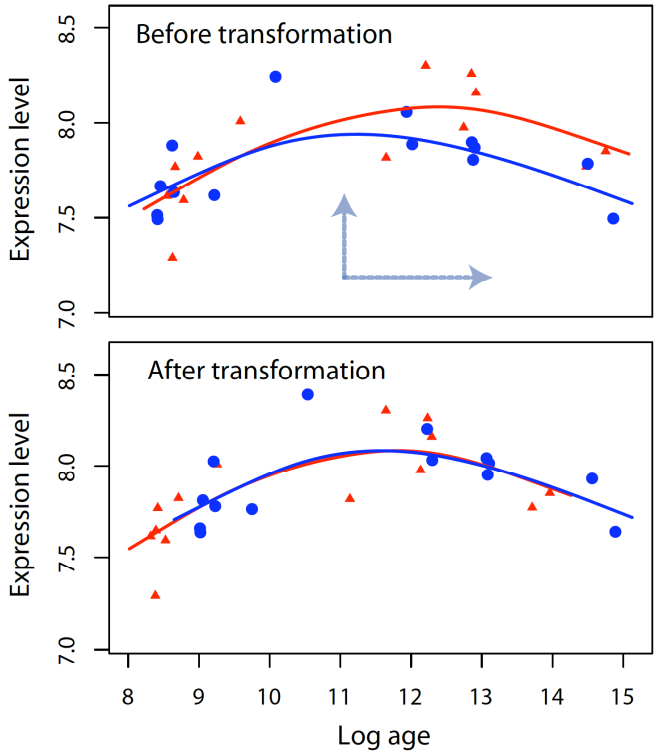

Fig. S3. How the heterochrony test works.

The *x*-axis represents log2 transformed conception ages in days, the *y*-axis, normalized and log2 transformed expression levels for a gene (UBE2B, ubiquitin-protein ligase B) chosen as an example from the dorsolateral prefrontal cortex dataset. **Upper panel:** Expression levels of the 14 humans (red triangles), the 14 chimpanzees (blue points). For this gene, the heterochrony test algorithm chooses an expression shift of magnitude +1.43, and an age-shift (C) of magnitude +0.82 (*i.e.* linear chimpanzee ages would be multiplied by $2^{0.82}$=1.8 to match the human curve). The arrows show the directions in which the chimpanzee curve would be shifted to match the human curve. **Lower panel:** The fit between the human and chimpanzee curves, after applying the expression and age shifts on the chimpanzee curve, *i.e.* adding 1.43 to the chimpanzee expression levels and 0.82 to chimpanzee log-transformed ages. The improvement in the fit is significant (F-test *p*-value = 0.011). Hence this gene is classified as a gene with delayed expression in humans compared to the chimpanzees.

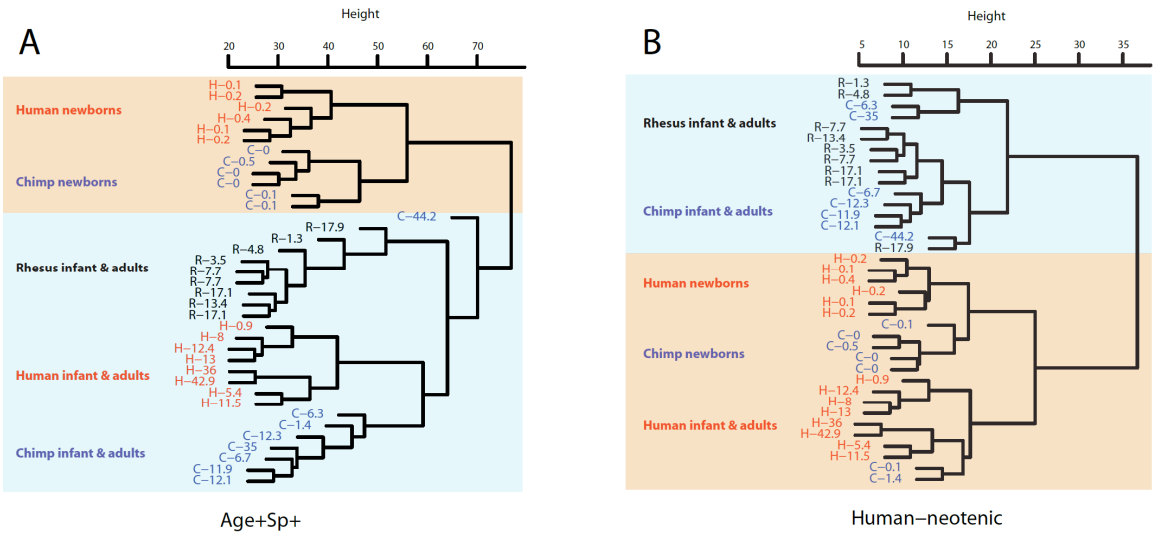

Fig. S4. Primate gene expression trees.

Hierarchical clustering trees of 14 humans, 14 chimpanzees, and 9 rhesus macaques, based on expression profiles of 1,225 genes which are age-related and differentially expressed between humans and chimpanzees (Age+,sp+) *(A)*, or 171 human-neotenic genes *(B)*. Genes are identified in the dorsolateral prefrontal cortex dataset, using relaxed significance cutoffs (p<0.1) for all tests. Individuals are indicated by the initial letter of their species: H – human, C – chimpanzee, R – rhesus macaque. The individuals' ages in years after birth is indicated after the dash. The trees were built based on Euclidean distances of normalized expression values per gene, and calculated using the "hclust" function in the R "stats" package.

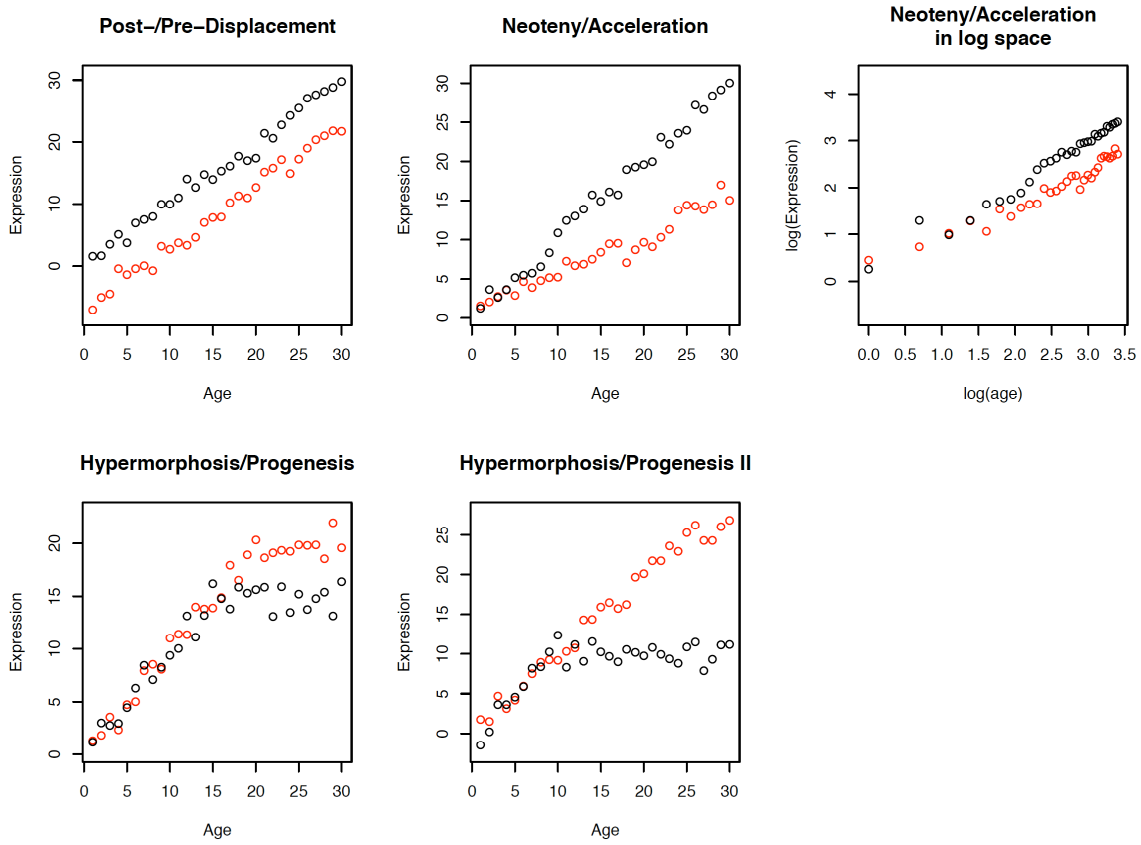

Fig. S5. Simulations of different modes of heterochrony.

The plots show simulated examples of three different modes of heterochrony, following the definitions in (29). In all plots, the red circles represent the reference species (*e.g.* human), which exhibits a delay in maturation in the form of either post-displacement, neoteny or hypermorphosis relative to the second species (*e.g.* chimpanzee), represented by black circles. The third plot on the upper panel is based on the same data as the previous one, except that age and expression level values are log transformed. The two plots in the lower panel represent two different simulated examples of hypermorphic development. Employing the heterochrony test and default criteria (using both the age-shift and expression-shift models), all five examples would be classified as showing "neoteny" in the reference species relative to the second species. Employing the heterochrony test and using only the age-shift model and *not* the expression-shift model ("condition set 3"), the examples of hypermorphosis would be classified as "accelerated". See text for description of criteria.

Supporting Tables

Table S1. Sample characteristics.

	Species	Brain region	Age days	Age (years /weeks)[a]	Error in age (years)[b]	Sex	Pmi[c]	pH	28S/18S[d]	Cause of death	Population origin
1	*Homo sapiens*	DLPFC	39	0.1	0	F	27	6.47	1.4	asphyxia	-
2	*Homo sapiens*	DLPFC	54	0.1	0	M	17	6.63	1.5	SIDS	-
3	*Homo sapiens*	DLPFC	56	0.2	0	M	11	6.86	n.a.	congenital heart defect	-
4	*Homo sapiens*	DLPFC	62	0.2	0	M	27	6.52	1.5	pneumonia	-
5	*Homo sapiens*	DLPFC	89	0.2	0	M	24	6.65	1.2	positional asphyxia	-
6	*Homo sapiens*	DLPFC	92	0.3	0	F	14	6.54	n.a.	SIDS	-
7	*Homo sapiens*	DLPFC	118	0.3	0	M	19	6.36	1.4	asphyxia	-
8	*Homo sapiens*	DLPFC	139	0.4	0	M	9	6.54	1.5	SIDS	-
9	*Homo sapiens*	DLPFC	141	0.4	0	M	5	6.81	1.6	asthma	-
10	*Homo sapiens*	DLPFC	188	0.5	0	F	22	6.82	1.7	SIDS	-
11	*Homo sapiens*	DLPFC	198	0.5	0	M	24	6.71	1.8	accident/ asphyxia	-
12	*Homo sapiens*	DLPFC	301	0.8	0	M	18	6.65	0.8	hypothermia	-
13	*Homo sapiens*	DLPFC	332	0.9	0	M	18	6.87	1.6	SUD	-
14	*Homo sapiens*	DLPFC	801	2.2	0	M	27	6.64	1.2	asthma	-
15	*Homo sapiens*	DLPFC	893	2.4	0	F	22	6.74	1.6	no anatomical cause	-
16	*Homo sapiens*	DLPFC	990	2.7	0	F	44	6.47	0.9	drowning	-
17	*Homo sapiens*	DLPFC	1692	4.6	0	M	18	6.92	1.3	accident	-
18	*Homo sapiens*	DLPFC	1773	4.9	0	M	19	6.74	n.a.	drowning	-
19	*Homo sapiens*	DLPFC	1969	5.4	0	M	17	6.74	1.7	drowning	-
20	*Homo sapiens*	DLPFC	2920	8	0	M	5	6.76	1.6	cardiac arhythmia	-
21	*Homo sapiens*	DLPFC	2920	8	0	F	20	6.78	0.9	asphyxia	-
22	*Homo sapiens*	DLPFC	4213	11.5	0	F	12	6.44	1.9	asthma	-

23	*Homo sapiens*	DLPFC	4534	12.4	0	M	16	6.82	1.7	drowning	-
24	*Homo sapiens*	DLPFC	4733	13	0	F	18	6.85	1.7	accident	-
25	*Homo sapiens*	DLPFC	6090	16.7	0	F	16	6.81	1.6	multiple injuries	-
26	*Homo sapiens*	DLPFC	6222	17	0	M	25	6.69	1.7	drowning	-
27	*Homo sapiens*	DLPFC	6456	17.7	0	M	16	6.83	1.5	accident	-
28	*Homo sapiens*	DLPFC	6505	17.8	0	M	12	6.8	1.5	accident/ asphyxia	-
29	*Homo sapiens*	DLPFC	7350	20.1	0	M	n.a.	n.a.	n.a.	n.a.	-
30	*Homo sapiens*	DLPFC	8006	21.9	0	M	13	6.96	1.6	MVA	-
31	*Homo sapiens*	DLPFC	8364	22.9	0	M	4	6.84	1.7	ASCVD	-
32	*Homo sapiens*	DLPFC	8623	23.6	0	F	14	6.57	1.4	asthma	-
33	*Homo sapiens*	DLPFC	9098	24.9	0	F	7	6.92	1.9	MVA	-
34	*Homo sapiens*	DLPFC	9262	25.4	0	F	16	6.73	1.8	accident	-
35	*Homo sapiens*	DLPFC	13138	36	0	M	13	6.73	1.5	coronary artery dis	-
36	*Homo sapiens*	DLPFC	14024	38.4	0	F	19	n.a.	1.2	n.a.	-
37	*Homo sapiens*	DLPFC	15672	42.9	0	M	18	6.49	1.1	accident	-
38	*Homo sapiens*	DLPFC	16855	46.2	0	M	18	6.75	1.5	accident	-
39	*Homo sapiens*	DLPFC	17317	47.4	0	M	12	6.56	1.6	ASCVD	-
40	*Pan troglodytes*	DLPFC	0	0	0	M	n.a.	n.a.	2.2	n.a.	Western
41	*Pan troglodytes*	DLPFC	1	0	0	F	n.a.	n.a.	1.6	n.a.	Central/ Eastern
42	*Pan troglodytes*	DLPFC	8	0	0	M	n.a.	n.a.	1.5	n.a.	Western
43	*Pan troglodytes*	DLPFC	40	0.1	0	M	n.a.	n.a.	1.4	n.a.	Western
44	*Pan troglodytes*	DLPFC	45	0.1	0	F	n.a.	n.a.	1.8	n.a.	Western
45	*Pan troglodytes*	DLPFC	186	0.5	±0.2	M	n.a.	n.a.	1.4	n.a.	Western
46	*Pan troglodytes*	DLPFC	525	1.4	0	F	n.a.	n.a.	1.6	n.a.	n.a.
47	*Pan troglodytes*	DLPFC	2313	6.3	0	F	n.a.	n.a.	1.4	n.a.	n.a.
48	*Pan troglodytes*	DLPFC	2447	6.7	0	M	n.a.	n.a.	2.4	n.a.	Western
49	*Pan troglodytes*	DLPFC	4361	11.9	0	M	n.a.	n.a.	1.6	n.a.	Western
50	*Pan troglodytes*	DLPFC	4415	12.1	0	M	n.a.	n.a.	3	n.a.	Western
51	*Pan troglodytes*	DLPFC	4480	12.3	0	M	n.a.	n.a.	1.7	n.a.	Western

52	*Pan troglodytes*	DLPFC	12784	35	±5	M	n.a.	n.a.	1.5	n.a.	n.a.
53	*Pan troglodytes*	DLPFC	16131	44.2	0	F	n.a.	n.a.	1.6	n.a.	Western
54	*Macaca mulatta*	DLPFC	467	1.3	0	F	~ 1	n.a.	1.2	n.a.	-
55	*Macaca mulatta*	DLPFC	1269	3.5	0	M	~ 1	n.a.	1.2	n.a.	-
56	*Macaca mulatta*	DLPFC	1760	4.8	0	M	~ 1	n.a.	1.4	n.a.	-
57	*Macaca mulatta*	DLPFC	2814	7.7	±0.5	M	~ 1	n.a.	1.4	n.a.	-
58	*Macaca mulatta*	DLPFC	2816	7.7	±0.5	M	~ 1	n.a.	1.4	n.a.	-
59	*Macaca mulatta*	DLPFC	4894	13.4	0	F	~ 1	n.a.	1.1	n.a.	-
60	*Macaca mulatta*	DLPFC	6230	17.1	±1	F	~ 1	n.a.	1.2	n.a.	-
61	*Macaca mulatta*	DLPFC	6230	17.1	±1	F	~ 1	n.a.	1.3	n.a.	-
62	*Macaca mulatta*	DLPFC	6524	17.9	0	F	~ 1	n.a.	1.2	n.a.	-
63	*Homo sapiens*	Caudate nucleus	128	0.4	0	M	27	6.66	2.1	n.a.	-
64	*Homo sapiens*	Caudate nucleus	141	0.4	0	M	5	6.81	2	asthma	-
65	*Homo sapiens*	Caudate nucleus	332	0.9	0	M	18	6.87	2	sud	-
66	*Homo sapiens*	Caudate nucleus	893	2.4	0	F	22	6.74	2	no anatomical cause	-
67	*Homo sapiens*	Caudate nucleus	4534	12.4	0	M	16	6.82	2	drowning	-
68	*Homo sapiens*	Caudate nucleus	4738	13	0	F	18	6.85	2	n.a.	-
69	*Homo sapiens*	Caudate nucleus	6222	17	0	M	25	6.69	2	drowning	-
70	*Homo sapiens*	Caudate nucleus	6384	17.5	0	M	16	6.67	2.1	n.a.	-
71	*Homo sapiens*	Caudate nucleus	8364	22.9	0	M	4	6.84	2	n.a.	-
72	*Homo sapiens*	Caudate nucleus	9098	24.9	0	M	7	6.92	2	MVA	-
73	*Homo sapiens*	Caudate nucleus	9253	25.4	0	F	16	6.73	2.1	n.a.	-
74	*Homo sapiens*	Caudate nucleus	14100	38.6	0	M	8	6.37	2.1	n.a.	-
75	*Homo sapiens*	Caudate nucleus	16855	46.2	0	M	18	6.75	2	accident	-
76	*Mus musculus*	Whole brain	0	0	0	M	~3 min.	n.a.	1.7	-	-
77	*Mus musculus*	Whole brain	0	0	0	M	~3 min.	n.a.	1.7	-	-
78	*Mus musculus*	Whole brain	0	0	0	M	~3 min.	n.a.	1.9	-	-
79	*Mus musculus*	Whole brain	14	2	0	M	~3 min.	n.a.	1.8	-	-
80	*Mus musculus*	Whole brain	14	2	0	M	~3 min.	n.a.	1.6	-	-

81	*Mus musculus*	Whole brain	14	2	0	M	~3 min.	n.a.	1.7	-	-
82	*Mus musculus*	Whole brain	56	8	0	M	~3 min.	n.a.	1.4	-	-
83	*Mus musculus*	Whole brain	56	8	0	M	~3 min.	n.a.	1.3	-	-
84	*Mus musculus*	Whole brain	56	8	0	M	~3 min.	n.a.	1.4	-	-
85	*Mus spretus*	Whole brain	0	0	0	M	~3 min.	n.a.	1.6	-	-
86	*Mus spretus*	Whole brain	0	0	0	M	~3 min.	n.a.	1.6	-	-
87	*Mus spretus*	Whole brain	0	0	0	M	~3 min.	n.a.	1.6	-	-
88	*Mus spretus*	Whole brain	14	2	0	M	~3 min.	n.a.	1.6	-	-
89	*Mus spretus*	Whole brain	14	2	0	M	~3 min.	n.a.	1.4	-	-
90	*Mus spretus*	Whole brain	14	2	0	M	~3 min.	n.a.	1.4	-	-
91	*Mus spretus*	Whole brain	56	8	0	M	~3 min.	n.a.	1.6	-	-
92	*Mus spretus*	Whole brain	56	8	0	M	~3 min.	n.a.	1.3	-	-
93	*Mus spretus*	Whole brain	56	8	0	M	~3 min.	n.a.	1.3	-	-
94	*Homo sapiens*	SFG	1	0.0	0	M	n.a.	n.a.	1.4	n.a.	-
95	*Homo sapiens*	SFG	4	0.0	0	M	n.a.	n.a.	1.5	n.a.	-
96	*Homo sapiens*	SFG	34	0.1	0	M	n.a.	n.a.	n.a.	n.a.	-
97	*Homo sapiens*	SFG	204	0.6	0	M	n.a.	n.a.	1.5	n.a.	-
98	*Homo sapiens*	SFG	2922	8.0	0	M	n.a.	n.a.	1.2	n.a.	-
99	*Homo sapiens*	SFG	5105	14.0	0	M	n.a.	n.a.	n.a.	n.a.	-
100	*Homo sapiens*	SFG	9277	25.4	0	M	n.a.	n.a.	1.4	n.a.	-
101	*Homo sapiens*	SFG	19457	53.3	0	M	n.a.	n.a.	1.5	n.a.	-
102	*Homo sapiens*	SFG	24090	66.0	±0.5	M	n.a.	n.a.	1.6	n.a.	-
103	*Pan troglodytes*	SFG	1	0.0	0	F	n.a.	n.a.	2.1	n.a.	Central/ Eastern
104	*Pan troglodytes*	SFG	8	0.0	0	M	n.a.	n.a.	1.5	n.a.	Western
105	*Pan troglodytes*	SFG	40	0.1	0	M	n.a.	n.a.	1.2	n.a.	Western
106	*Pan troglodytes*	SFG	525	1.4	0	F	n.a.	n.a.	1.2	n.a.	n.a.
107	*Pan troglodytes*	SFG	2313	6.3	0	F	n.a.	n.a.	1.2	n.a.	n.a.
108	*Pan troglodytes*	SFG	4415	12.1	0	M	n.a.	n.a.	2	n.a.	Western
109	*Pan troglodytes*	SFG	4480	12.3	0	M	n.a.	n.a.	1.6	n.a.	Western

110	*Pan troglodytes*	SFG	12784	35	±5	M	n.a.	n.a.	1.6	n.a.	n.a.
111	*Pan troglodytes*	SFG	16131	44.2	0	F	n.a.	n.a.	1.5	n.a.	Western
112	*Macaca mulatta*	SFG	16	0.0	0	M	~ 1	n.a.	1.2	n.a.	-
113	*Macaca mulatta*	SFG	20	0.1	0	M	~ 1	n.a.	1.2	n.a.	-
114	*Macaca mulatta*	SFG	153	0.4	0	M	~ 1	n.a.	1.4	n.a.	-
115	*Macaca mulatta*	SFG	310	0.8	0	M	~ 1	n.a.	1.4	n.a.	-
116	*Macaca mulatta*	SFG	3389	9.3	0	M	~ 1	n.a.	1.4	n.a.	-
117	*Macaca mulatta*	SFG	7391	20.2	0	M	~ 1	n.a.	1.1	n.a.	-
118	*Macaca mulatta*	SFG	8104	22.2	0	M	~ 1	n.a.	1.2	n.a.	-
119	*Macaca mulatta*	SFG	9518	26.1	0	M	~ 1	n.a.	1.3	n.a.	-
120	*Macaca mulatta*	SFG	10220	28.0	±1	F	~ 1	n.a.	1.	n.a.	-

Abbreviations: n.a.: Data not available; DLPFC: Dorsolateral prefrontal cortex; SFG: Superior frontal gyrus; SIDS: sudden infant death; SUD: sudden death; MVA: motor vehicle accident; ASCVD: arteriosclerotic cardiovascular disease.

[a] Age from birth, in years for the primate and in weeks for the mouse species.

[b] Possible error in age estimates in years for chimpanzee and rhesus individuals for which the birth date was not exactly know.

[c] Post-mortem interval in hours.

[d] 28S/18S rRNA ratio, indicating RNA quality.

Table S2. Expression variance explained by age, species identity and sex.

The table shows the proportion of total expression variation (R^2) explained by age, species identity, or sex in three brain datasets. SI section *"The influence of age, species identity and sex on expression - a general view"* describes how the proportion is calculated. For the effects of these factors tested gene by gene, see Table S12, Table 15, and Table S16.

		$\%R^2_f$ [a]	*p*-value [b]	%FDR [c]
Human-chimpanzee dorsolateral PFC 39 + 14	**Age**	28.73	<0.001	19
	Species	16.77	<0.001	10
	Sex	1.18	0.90	149
	Sex-corrected[d]	1.82	0.54	101
Human-chimpanzee superior frontal gyrus 9 + 9	**Age**	33.60	<0.001	52
	Species	16.22	<0.001	33
***M.musculus* - *M.spretus* 9 + 9**	**Age**	53.32	<0.001	19
	Species	15.14	0.02	33

[a] Mean percent of variance explained by each factor among all expressed genes.
[b] *p*-value for R^2_f being larger than random based on 1,000 permutations.
[c] Percent false discovery rate based on permutations; see text.
[d] The effect of sex after factoring out the effects of age and species; see SI section *"The influence of age, species identity and sex on expression"* for details.

Table S3. Biological processes overrepresented among age-related genes.

The Gene Ontology (GO) biological process groups that are overrepresented among age-related genes with GO annotation (n=4,349) compared to non-age-related genes (n=1,767). Age-related genes were identified among the 39 humans of the dorsolateral prefrontal cortex dataset, at false discovery rate 10%.

	Gene Ontology group	Age-related	Other	*p*-value [a]
GO:0007154	Cell communication	1362	456	9.5E-06
GO:0032501	Multicellular organismal process	1112	367	3.5E-05
GO:0048666	Neuron development	99	15	5.1E-05
GO:0031175	Neurite development	90	13	6.6E-05
GO:0007267	Cell-cell signaling	215	50	9.0E-05
GO:0050877	Neurological system process	262	66	1.3E-04
GO:0000904	Cellular morphogenesis during differentiation	79	11	1.3E-04
GO:0048468	Cell development	218	53	2.2E-04
GO:0048667	Neuron morphogenesis during differentiation	75	11	3.1E-04
GO:0048812	Neurite morphogenesis	75	11	3.1E-04
GO:0030182	Neuron differentiation	122	25	5.4E-04
GO:0019226	Transmission of nerve impulse	156	36	7.3E-04
GO:0007165	Signal transduction	1220	426	8.5E-04
GO:0007409	Axonogenesis	70	11	9.0E-04
GO:0022008	Neurogenesis	153	36	1.1E-03
GO:0007399	Nervous system development	345	101	1.2E-03
GO:0007268	Synaptic transmission	139	32	1.4E-03
GO:0048856	Anatomical structure development	736	245	1.6E-03
GO:0048699	Generation of neurons	140	33	1.9E-03
GO:0048731	System development	644	213	2.5E-03
GO:0003008	System process	317	95	3.5E-03
GO:0007155	Cell adhesion	294	87	3.6E-03
GO:0022610	Biological adhesion	294	87	3.6E-03
GO:0000902	Cell morphogenesis	158	41	4.5E-03
GO:0032989	Cellular structure morphogenesis	158	41	4.5E-03
GO:0007275	Multicellular organismal development	813	282	6.0E-03
GO:0030030	Cell projection organization and biogenesis	122	31	9.2E-03
GO:0032990	Cell part morphogenesis	122	31	9.2E-03
GO:0048858	Cell projection morphogenesis	122	31	9.2E-03
GO:0007166	Cell surface receptor linked signal transduction	512	171	9.7E-03

[a] Fisher's exact test *p*-value (one-sided) for a higher number of human-neotenic genes in a GO group compared to genes in the other phylo-ontogenetic categories.

Table S4. Correlation between age-related expression changes in primates and mice.

The table shows the median correlation coefficients among 2,599 age+ genes (lower triangle) and among 388 age- genes (upper triangle) between the indicated species. For each human-mouse ortholog, we calculated the Pearson correlation coefficient between each pair of species using mean expression levels in 3 age groups: newborns, adolescents/pups, and adults (see SI section *"Conservation of age-related changes in mice and primates"* and Fig. S2).

	Human	**Chimpanzee**	***M.musculus***	***M.spretus***
Human	-	0.41	-0.02	0.07
Chimpanzee	0.90	-	0.10	0.23
M.musculus	0.83	0.76	-	0.53
M.spretus	0.83	0.79	0.98	-

Table S5. Comparison of prefrontal cortex with caudate nucleus in developmental timing.

The table shows the number of genes classified in different categories and the excess of genes showing delayed dorsolateral prefrontal cortex development (cortex-neotenic genes) with respect to the caudate nucleus.

	39 cortex *vs* 13 caudate nucleus [a]	**13 cortex *vs* 13 caudate nucleus** [b]
Age-related and differentially expressed (FDR=10%)	5484	2979
Cortex-neotenic [c]	2494	1310
Cortex-accelerated [c]	1356	951
% Cortex-neotenic	64.8	57.9
p neoteny[d]	$<10^{-16}$	$<10^{-16}$

[a] Analyses conducted using all 39 human dorsolateral prefrontal cortex samples.
[b] Analyses conducted using only 13 human dorsolateral prefrontal cortex samples with ages matched to the 13 caudate nucleus samples (Fig. S1D).
[c] The number of genes that are age-related and differentially expressed between regions, and are also classified as cortex-neotenic or cortex-accelerated relative to the caudate nucleus.
[d] The one-sided binomial test *p*-value for the frequency of cortex-neotenic genes among all heterochronic genes >50%.

Table S6. Phylo-ontogenetic classification under different assumptions and criteria.
(continued on next pages)

The table shows the number of genes classified in phylo-ontogenetic categories and the excess of human-neotenic genes under different assumptions or criteria. In all cases we have used the dorsolateral prefrontal cortex dataset and compared 14 humans with 14 chimpanzees (unless otherwise indicated). Gene sets identified under different criteria (here and in Table 1) overlap with the main result by at least 40%.

	Assumptions/ Criteria/ Subset	Main result (dorsolateral pfc)	Expressed in 100% [a]	Mean > 50% [b]	No RNA/pH effect [c]	Only males [d]
	Number of expressed genes	3075	2073	2237	1159	3075
FDR = 10% for age and species effects; $p<0.05$ for heterochrony and lineage (one-sided) tests	Human-neotenic	114	85	93	22	139
	Human-accelerated	65	48	54	12	88
	Chimpanzee-accelerated	46	37	39	9	77
	Chimpanzee-neotenic	74	57	57	12	106
	p neoteny [m]	2.E-04	8.E-04	8.E-04	6.E-02	4.E-04
	p human specificity [n]	4.E-08	8.E-06	1.E-06	1.E-02	1.E-05
Relaxed cutoffs: $p < 0.10$ in all four tests	Human-neotenic	171	132	138	35	226
	Human-accelerated	100	66	77	27	135
	Chimpanzee-accelerated	81	62	65	18	143
	Chimpanzee-neotenic	121	95	98	31	187
	p neoteny [m]	1.E-05	2.E-06	2.E-05	2.E-01	1.E-06
	p human specificity [n]	7.E-09	3.E-07	2.E-07	1.E-02	9.E-06

[a] Using genes expressed in *all* 39 humans and 14 chimpanzees (instead of genes expressed in >1/3 of individuals in either species, as in the main analysis).
[b] Using genes with mean log expression level above the mean expression of 50% of all genes on the array among all humans and chimpanzees (instead of using MAS5 detection cutoffs, as in the main analysis).
[c] Using genes which do not show pH or RNA quality effects, or any interactions of these factors with species or age in their expression profiles (at $p<0.05$) (see Table S12).
[d] Using only males: 26 humans *vs.* 9 chimpanzees.
[m] The one-sided binomial test *p*-value for neoteny: More than 50% of genes that are assigned to the human lineage show delayed development in human relative to chimpanzee.
[n] The one-sided binomial test *p*-value for human-specificity: More than 50% of genes that show delayed development in human vs. chimpanzee are assigned to the human lineage.
Note: Repeating the analysis by randomly choosing 8 probes within each probeset with replacement, the *p*-values for both human-specificity and neoteny were significant at $p<0.05$ in 91/100 bootstrap trials (97/100 using relaxed cutoffs).

Table S6 continued.

	Assumptions/ Criteria/ Subset	Main result (dorsolateral pfc)	Linear age scale [e]	Alternative lineage test [f]	Chimps as reference [g]	Neoteny condition 2 [h]	Neoteny condition 3 [i]
	Number of expressed genes	3075	3075	3075	3075	3075	3075
FDR = 10% for age and species effects; $p<0.05$ for heterochrony and lineage (one-sided) tests	Human-neotenic	114	66	86	126	51	102
	Human-accelerated	65	19	48	73	0	84
	Chimpanzee-accelerated	46	21	33	59	9	39
	Chimpanzee-neotenic	74	42	44	78	1	82
	p neoteny [m]	2.E-04	2.E-07	7.E-04	1.E-04	4.E-16	1.E-01
	p human specificity [n]	4.E-08	7.E-07	6.E-07	5.E-07	2.E-08	6.E-08
Relaxed cutoffs: $p < 0.10$ in all four tests	Human-neotenic	171	133	120	198	186	152
	Human-accelerated	100	67	74	119	6	123
	Chimpanzee-accelerated	81	58	55	112	42	67
	Chimpanzee-neotenic	121	114	65	138	9	136
	p neoteny [m]	1.E-05	2.E-06	6.E-04	5.E-06	1.E-47	5.E-02
	p human specificity [n]	7.E-09	3.E-08	5.E-07	6.E-07	4.E-23	5.E-09

[e] Using the linear age scale rather than log transformation throughout the analyses.
[f] Using a multiple regression-based test for lineage assignment: Genes differently expressed between human and rhesus, but not rhesus and human, are considered human-specific, and vice versa (see SI text).
[g] Using chimpanzee as the reference species, instead of human.
[h] Using more stringent conditions to define neoteny: Requiring that the age-shift model (model A) is always significantly better than the expression-shift model (see Supporting Note 2).
[i] Using only the age-shift model (model A) of the heterochrony test, ignoring a possible expression shift (see SI text).
[m] The one-sided binomial test *p*-value for neoteny: More than 50% of genes that are assigned to the human lineage show delayed development in human relative to chimpanzee.
[n] The one-sided binomial test *p*-value for human-specificity: More than 50% of genes that show delayed development in human vs. chimpanzee are assigned to the human lineage.

Table S6 continued.

	Assumptions/ Criteria/ Subset	Main result (dorsolateral pfc)	Using all humans [j]	Gestation time diff = 20 days [k]	Gestation time diff = 20 days [l]
	Number of expressed genes	3075	3075	3075	3075
FDR = 10% for age and species effects; $p<0.05$ for heterochrony and lineage (one-sided) tests	Human-neotenic	114	154	121	131
	Human-accelerated	65	88	59	54
	Chimpanzee-accelerated	46	86	59	63
	Chimpanzee-neotenic	74	118	78	72
	p neoteny [m]	2.E-04	1.E-05	2.E-06	7.E-09
	p human specificity [n]	4.E-08	7.E-06	2.E-06	6.E-07
Relaxed cutoffs: $p < 0.10$ in all four tests	Human-neotenic	171	255	208	232
	Human-accelerated	100	138	107	97
	Chimpanzee-accelerated	81	148	97	110
	Chimpanzee-neotenic	121	193	129	122
	p neoteny [m]	1.E-05	2.E-09	7.E-09	3.E-14
	p human specificity [n]	7.E-09	5.E-08	1.E-10	2.E-11

[j] Using all 39 humans in the age, species, lineage and heterochrony tests (instead of 14 age-matched humans).
[k] Assuming a 20 day shorter gestation time in chimpanzees than humans (instead using 280 days for both species, as in the main analysis).
[l] Assuming a 51 day shorter gestation time in chimpanzees than humans (instead using 280 days for both species, as in the main analysis).
[m] The one-sided binomial test *p*-value for neoteny: More than 50% of genes that are assigned to the human lineage show delayed development in human relative to chimpanzee.
[n] The one-sided binomial test *p*-value for human-specificity: More than 50% of genes that show delayed development in human vs. chimpanzee are assigned to the human lineage.

Table S7. The false negative and false positive rates of phylo-ontogenetic classification.

In order to estimate the false positive rates of the tests for age, species, lineage and heterochrony, we compare 14 humans to 14 other humans in the dorsolateral prefrontal cortex dataset. In order to estimate the false negative rates, we simulate slower or faster developmental rates in the reference group relative to the other, by multiplying individuals' ages by 1/2 or 2. For the tests for age and species effects, we use the same *p*-value cutoffs as in the original 14 human subset. The total number of age-related genes, after applying the human-chimp-rhesus mask, is 1,527 (at FDR=10%) and 1,979 (at $p<0.10$). We estimate the false positive rate as ~10%, and the false negative rate ~75% (~30% and ~65% using relaxed significance cutoffs, respectively).

	Groups compared	**14 humans *vs.* 14 chimps**	**14 humans *vs.* 14 humans I** [a]	**14 humans *vs.* 14 humans II** [a]	**14 humans *vs.* 14 humans**	**14 humans *vs.* 14 humans**
	Simulated difference in developmental rate	-	0	0	0.5X	2X
FDR = 10% for age and species effects; $p<0.05$ for heterochrony and lineage (one-sided) tests	Human-neotenic	114	13	8	204	1
	Human-accelerated	65	1	7	0	99
	Chimpanzee-accelerated	46	6	0	208	5
	Chimpanzee-neotenic	74	10	8	2	399
	p neoteny [b]	2.E-04	9.E-04	5.E-01	4.E-62	1.E+00
	p human specificity [c]	4.E-08	8.E-02	4.E-03	6.E-01	1.E+00
Relaxed cutoffs: $p < 0.10$ in all four tests	Human-neotenic	171	56	33	339	5
	Human-accelerated	100	15	29	0	204
	Chimpanzee-accelerated	81	40	7	361	11
	Chimpanzee-neotenic	121	30	40	2	595
	p neoteny [b]	1.E-05	5.E-07	4.E-01	9.E-103	1.E+00
	p human specificity [c]	7.E-09	6.E-02	2.E-05	8.E-01	1.E+00

[a] The same test conducted on the same two sets of 14 humans, but switching the reference group in the heterochrony test.
[b] The binomial test *p*-value for neoteny: More than 50% of genes that are assigned to the human lineage show delayed development in human relative to chimpanzee (or reference group).
[c] The binomial test *p*-value for human-specificity: More than 50% of genes that show delayed development in human vs. chimpanzee (or reference group) are assigned to the human lineage.

Table S8. Correspondence between datasets in age, species, lineage, and heterochrony tests.

We compare different datasets in terms of how many genes are consistently identified in the same gene sets. The comparisons are: (1) the dorsolateral prefrontal cortex dataset versus the superior frontal gyrus dataset, (2) humans vs chimpanzees, (3) humans vs mice, (4) the dorsolateral prefrontal cortex dataset vs a previously published human-chimpanzee dataset (1). The gene sets are: age-related or not (Age+/Age-); differentially expressed between species or not (Sp+/Sp-); assigned to the human or to the chimpanzee lineage (Human/Chimp specific); showing delayed or accelerated development in human relative to chimpanzee (Neotenic/Accelerated); assigned to the human-neotenic category or one of the other three phylo-ontogenetic categories (Human-neotenic/Other).

	Dataset 1	**Dorsolateral PFC** [a]						**Human** [c]	**Human** [c]	**Dorso-lateral PFC** [f]
	Dataset 2	**Superior frontal gyrus** [b]						**Chimp** [d]	**Mouse** [e]	**PFC - 2** [g]
	Gene set 1 [h]	**Age+**	**Sp+**	**Age+ & Sp+**	**Human specific**	**Neo-tenic**	**Human neo-tenic**	**Age+**	**Age+**	**Sp+**
	Gene set 2 [h]	**Age-**	**Sp-**	**Age+ & Sp+**	**Chimp specific**	**Acceler ated**	**Other**	**Age-**	**Age-**	**Sp-**
FDR = 10% for age and species effects; $p<0.05$ for heterochrony and lineage (one-sided) tests	GS1 & GS1	640	442	112	16	38	6	1087	829	1216
	GS1 & GS2	180	360	102	5	13	2	440	184	1104
	GS2 & GS1	798	881	161	7	2	1	470	700	1207
	GS2 & GS2	1245	1180	265	12	8	10	1078	285	3436
	Odd's ratio	5.5	1.6	1.8	5.2	11.1	22.9	5.7	1.8	3.1
	p overlap [i]	8.E-83	2.E-09	3.E-04	1.E-02	2.E-03	6.E-03	1.E-116	8.E-09	1.E-103
Relaxed cutoffs $p < 0.10$ in all four tests	GS1 & GS1	1296	1021	464	92	97	29	1539	1057	1635
	GS1 & GS2	384	545	284	29	41	15	440	242	1275
	GS2 & GS1	565	742	304	44	8	7	515	532	1328
	GS2 & GS2	618	555	244	61	35	40	581	167	2725
	Odd's ratio	3.7	1.4	1.3	4.4	10.2	10.7	3.9	1.4	2.6
	p overlap [i]	3.E-59	7.E-06	1.E-02	1.E-07	2.E-09	6.E-07	2.E-66	3.E-03	6.E-85

[a] The dorsolateral PFC dataset with 14 humans and 14 chimps, and using the human-chimp-rhesus mask.

[b] The superior frontal gyrus dataset with 9 humans and 9 chimps, and using the human-chimp-rhesus mask.
[c] The dorsolateral PFC dataset with 14 humans, and using the human-chimp-rhesus mask.
[d] The dorsolateral PFC dataset with 14 chimps, and using the human-chimp-rhesus mask.
[e] 9 Mus musculus individuals in the mouse datasets, using the musculus-spretus mask.
[f] The dorsolateral PFC dataset with 6 human and 5 chimps adults, and using the human-chimp mask. The overlap is similar (odd's ratio=2.9) when we use all 39 humans and 14 chimps.
[g] The dorsolateral PFC dataset from (1), containing 6 humans and 5 chimps, and using the human-chimp mask.
[h] The gene sets as indicated in the table description. For example, in the first column, the two gene sets are age-related and not age-related genes. GS1 & GS1 in the first column indicates the number of genes that are age-related in both the DLPFC and SFG datasets; GS1 & GS2 indicates genes that are classified as age-related in the DLPFC dataset, but not age-related in the SFG dataset; GS2 & GS1 indicates genes that are classified as not age-related in the DLPFC dataset, but age-related in the SFG dataset.
[i] The Fisher's exact test *p*-value (one-sided) for a larger than randomly expected overlap (odd's ratio > 1) between gene sets identified in the two datasets.

Table S9. Overlap between human-neotenic genes and grey- and white-matter-specific genes.

We compare human-neotenic genes with genes in the three other phylo-genetic categories with respect to the overlap with grey- or white-matter specific genes (n=1,155 and n=578, respectively), compared to all 9,892 genes expressed in a previously published grey/white matter experiment (2).

	Gene sets [a]	**Dorsolateral PFC** [b]	**Superior Frontal Gyrus** [c]
FDR = 10% for age and species effects; $p<0.05$ for heterochrony and lineage (one-sided) tests	Grey & human-neotenic	21	38
	Grey & other cat.	25	20
	Non-grey & human-neotenic	94	151
	Non-grey & other cat.	137	137
	Odd's ratio [d]	1.2	1.7
	p overlap [d]	3.E-01	5.E-02
Relaxed cutoffs $p < 0.10$ in all four tests	Grey & human-neotenic	36	89
	Grey & other cat.	39	51
	Non-grey & human-neotenic	139	427
	Non-grey & other cat.	230	486
	Odd's ratio [d]	1.5	2.0
	p overlap [d]	6.E-02	1.E-04

FDR = 10% for age and species effects; $p<0.05$ for heterochrony and lineage (one-sided) tests	White & human-neotenic	4	9
	White & other cat.	9	8
	Non-white & human-neotenic	94	151
	Non- white & other cat.	137	137
	Odd's ratio [d]	0.6	1.0
	p overlap [d]	8.E-01	6.E-01
Relaxed cutoffs $p < 0.10$ in all four tests	White & human-neotenic	5	11
	White & other cat.	17	11
	Non-white & human-neotenic	139	171
	Non- white & other cat.	230	163
	Odd's ratio [d]	0.5	1.0
	p overlap [d]	1.E+00	6.E-01

[a] Grey: grey-matter specific genes. Non-grey: all other genes. Other cat.: The 3 phylo-ontogenetic except human-neotenic genes.
[b] The dorsolateral PFC dataset with 14 humans and 14 chimps, and using the human-chimp-rhesus mask.
[c] The superior frontal gyrus dataset with 9 humans and 9 chimps, and using the human-chimp-rhesus mask.
[d] The odd's ratio for the overlap between grey- or white-matter specific genes and human-neotenic genes, and the Fisher's exact test *p*-value (two-sided) for a non-random overlap (odd's ratio $\neq 1$) between gene sets identified in the two datasets.

Table S10. Biological processes overrepresented among human-neotenic genes in the dorsolateral PFC.

The Gene Ontology (GO) biological process groups that are overrepresented among human-neotenic genes (n=141) compared to genes in the other phylo-ontogenetic categories (n=250). We use relaxed significance cutoffs to define phylo-ontogenetic gene sets (at $p<0.1$) for all four tests. The significance for overall enrichment in biological process categories is only $p=0.24$ – therefore the results should be treated with caution. The GO groups shared with the superior frontal gyrus dataset are in bold.

	Gene Ontology group	**Human-neotenic**	**Other**	***p*-value [a]**
GO:0032502	Developmental process	53	59	0.003
GO:0007275	Multicellular organismal development	38	40	0.007
GO:0016032	Viral reproduction	6	1	0.01
GO:0015674	Di-, tri-valent inorganic cation transport	7	2	0.013
GO:0007010	Cytoskeleton organization and biogenesis	12	7	0.013
GO:0000003	Reproduction	11	6	0.014
GO:0001666	Response to hypoxia	4	0	0.016
GO:0019058	Viral infectious cycle	4	0	0.016
GO:0051179	Localization	47	58	0.021
GO:0022415	Viral reproductive process	5	1	0.025
GO:0006810	Transport	42	51	0.025
GO:0009888	Tissue development	9	5	0.027
GO:0006816	Calcium ion transport	6	2	0.028
GO:0006928	Cell motility	6	2	0.028
GO:0007517	Muscle development	6	2	0.028
GO:0008632	**Apoptotic program**	6	2	0.028
GO:0051674	Localization of cell	6	2	0.028
GO:0022414	Reproductive process	7	3	0.029
GO:0030001	Metal ion transport	12	9	0.035
GO:0051234	Establishment of localization	42	53	0.039
GO:0000041	Transition metal ion transport	3	0	0.046
GO:0007346	Regulation of mitotic cell cycle	3	0	0.046
GO:0007601	Visual perception	3	0	0.046
GO:0008633	Activation of pro-apoptotic gene products	3	0	0.046
GO:0019748	Secondary metabolic process	3	0	0.046
GO:0050953	Sensory perception of light stimulus	3	0	0.046
GO:0050877	**Neurological system process**	13	11	0.048
GO:0032501	Multicellular organismal process	45	59	0.048

[a] Fisher's exact test *p*-value (one-sided) for a higher number of human-neotenic genes in a GO group compared to genes in the other phylo-ontogenetic categories.

Table S11. Biological processes overrepresented among human-neotenic genes in the superior frontal gyrus.

The Gene Ontology (GO) biological process groups that are overrepresented among human-neotenic genes (n=516) compared to genes in the other phylo-ontogenetic categories (n=560). We use relaxed significance cutoffs to define phylo-ontogenetic gene sets (at $p<0.1$) for all four tests. The significance for overall enrichment in biological process categories is only $p=0.31$ – therefore the results should be treated with caution. The GO groups shared with the dorsolateral prefrontal cortex dataset are in bold.

	Gene Ontology group	Human-neotenic	Other	*p*-value [a]
GO:0000165	MAPKKK cascade	13	2	0.002
GO:0016485	Protein processing	8	0	0.003
GO:0008632	**Apoptotic program**	10	1	0.004
GO:0006350	Transcription	115	94	0.014
GO:0008361	Regulation of cell size	8	1	0.014
GO:0016049	Cell growth	8	1	0.014
GO:0043284	Biopolymer biosynthetic process	146	125	0.014
GO:0045449	Regulation of transcription	108	90	0.024
GO:0009880	Embryonic pattern specification	5	0	0.025
GO:0016540	Protein autoprocessing	5	0	0.025
GO:0046777	Protein amino acid autophosphorylation	5	0	0.025
GO:0009987	Cellular process	468	486	0.027
GO:0006355	Regulation of transcription, DNA-dependent	103	86	0.029
GO:0051252	Regulation of RNA metabolic process	104	87	0.029
GO:0032774	RNA biosynthetic process	107	90	0.029
GO:0008284	Positive regulation of cell proliferation	13	5	0.032
GO:0006351	Transcription, DNA-dependent	106	90	0.034
GO:0019219	Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	111	95	0.035
GO:0010558	Negative regulation of macromolecule biosynthetic process	18	9	0.037
GO:0048519	Negative regulation of biological process	59	45	0.037
GO:0050789	Regulation of biological process	261	252	0.038
GO:0050877	**Neurological system process**	30	19	0.039
GO:0010556	Regulation of macromolecule biosynthetic process	112	97	0.041
GO:0042127	Regulation of cell proliferation	25	15	0.043
GO:0050794	Regulation of cellular process	253	245	0.047
GO:0010468	Regulation of gene expression	115	101	0.048
GO:0009889	Regulation of biosynthetic process	112	98	0.048
GO:0043406	Positive regulation of MAP kinase activity	6	1	0.050
GO:0040007	Growth	12	5	0.050

[a] Fisher's exact test *p*-value (one-sided) for a higher number of human-neotenic genes in a GO group compared to genes in the other phylo-ontogenetic categories.

Table S12. Effects of biological and technical factors on gene expression.

The table shows the effects of age, species difference, sex, RNA quality and brain pH, and interactions between these factors, on gene expression in the brain. Effects were measured using linear regression models on all 7,958 expressed genes in the human-chimpanzee dorsolateral prefrontal cortex dataset. The significance of each factor is assessed based on the number of genes with F-test $p<0.05$ for that factor, compared to 1,000 random permutations. The false discovery rate (FDR) is also estimated by these permutations. Factors separated by colons indicate interaction effects.

Main factor		**% significant among all expressed**	***p*-value**	**%FDR**
Species (Sp)	Sp	60.5	**<0.001**	7
	Sex:Sp	6.5	0.236	63
	Age:Sp	23.1	**0.006**	20
	Sex:Age:Sp	11.5	0.153	43
Age	Age	67.3	**<0.001**	10
	Sex:Age	6.8	0.252	63
	Age:Sp	23.1	**0.072**	31
	Sex:Age:Sp	11.5	0.360	79
Sex	Sex	4.4	0.950	279
	Sex:Age	6.8	0.339	82
	Sex:Sp	6.5	0.772	138
	Sex:Age:Sp	11.5	0.338	80
RNA quality (28S/18S)	Rna	32.8	**0.039**	40
	Rna:Age	6.2	0.394	85
	Rna:Sp	10.2	0.347	80
	Rna:Age:Sp	9.1	0.491	99
Brain pH	pH	26.7	**0.045**	33
	pH:Age	3.6	0.706	129

Table S13. Correlation between age and clinical and demographic variables.

Results are based on the 39 human dorsolateral prefrontal cortex samples.

	N [a]	Mini-mum [a]	Median [a]	Maxi-mum [a]	*r* [b]	*P* [b]	rho [c]	*P* [c]
28S/18S [d]	35	0.8	1.5	1.9	0.09	0.61	0.17	0.33
5'-3' slope [e]	39	3.63	4.15	5.24	-0.06	0.72	-0.27	0.10
pH [f]	37	6.36	6.74	6.96	0.09	0.61	0.25	0.13
PMI [g]	38	4	17.5	44	-0.27	0.10	-0.32	0.05
Sex [h]	39	-	-	-	0.00	0.98	-0.07	0.67

[a] Number of samples with estimates for a variable, and the minimum, median and maximum values of the variable among the samples.
[b] Pearson coefficient and *p*-value for a parametric correlation between age and the variable in question.
[c] Spearman coefficient and *p*-value for a non-parametric correlation between age and the variable in question.
[d] 28S/16S RNA ratio.
[e] The slope between expression intensity and 5' to 3' position of probes in a probeset, averaged across all probesets, calculated by the "AffyRNADeg" function (R "affy" package). To calculate this measure, we used the original Affymetrix 133Plus2.0 CDF file.
[f] Brain pH.
[g] Post-mortem interval: hours between subject death and tissue storage.
[h] Sex of individuals; females coded as 0 and males as 1.

Table S14. The number of probes and genes in the microarray datasets used in the study.

We masked probes that do not match the respective species' genomes perfectly and at a unique location. We only accepted genes with at least 8 probes. Genes which are detected among 1/3 of subjects in at least one species are considered expressed. Genes are synonymous to probe sets where Custom CDF files (3) are used. See SI text for details.

Species	Brain region	N individuals	CDF	Probes before mask	Probes after mask	Genes before mask [a]	Genes after mask, with ≥ 8 probes [a]	Probes per gene (mean) [a]	Probes per gene (median) [a]	Expressed genes [b]
Human	Dorso-lateral PFC + Caudate nucleus	39+13	U133P2 Ensembl Custom CDF	276826	-	17427	15964	17	11	10543
Human + chimpanzee	Dorso-lateral PFC	39+14	U133P2 H-C Masked Custom CDF	276826	208941	17427	11956	15	12	7958
Human + chimpanzee + rhesus	Dorso-lateral PFC	39+14 +9	U133P2 H-C-R Masked Custom CDF	276826	86559	17427	3747	13	11	3075
Human + chimpanzee + rhesus	Superior Frontal Gyrus	9+ 9+ 9	HuGene1.0 H-C-R Masked Affymetrix CDF	862560	235751	33843	20882	16	13	11333
Mus musculus + *Mus spretus*	Whole Brain	9+ 9	MG430_2 M-S Masked Custom CDF	239706	194078	16316	12346	14	11	8362

[a] In this column, "gene" refers to Ensembl genes in datasets where we use Custom CDF files. For the HuGene 1.0 dataset, "gene" refers to transcripts as defined by Affymetrix.
[b] In this column, "gene" refers to Ensembl genes in all datasets.

Table S15. The proportion of age-related genes.

The false discovery rate (FDR) is estimated by 1,000 permutations. The permutation based *p*-value for the age effect is < 0.005 in all datasets. We choose a *p*-value cutoff for each dataset/subset that ensures an FDR close to 10%, shown in bold.

	***p*-value cutoff**	**Number of significant genes**	**% significant among all expressed**	**% FDR**
Dorsolateral PFC 39 humans	0.05	5740	72.1	11.1
	0.045	**5653**	**71.0**	**10.1**
	0.01	4699	59.0	2.4
	0.005	4350	54.7	1.2
	0.001	3669	46.1	0.2
Dorsolateral PFC 14 humans	0.05	4227	53.1	14.8
	0.03	**3713**	**46.7**	**9.8**
	0.01	2842	35.7	4.1
	0.005	2409	30.3	2.4
	0.001	1566	19.7	0.7
Dorsolateral PFC 14 chimpanzees	0.05	4399	55.3	15.2
	0.03	**3857**	**48.5**	**10.2**
	0.01	2901	36.5	4.3
	0.005	2403	30.2	2.5
	0.001	1534	19.3	0.8
S Frontal Gyrus 9 humans	0.05	4558	40.2	20.2
	0.015	**2690**	**23.7**	**9.6**
	0.01	2228	19.7	7.7
	0.005	1625	14.3	5.1
	0.001	733	6.5	2.3
Mouse 9 *Mus musculus*	**0.075**	**6106**	**73.0**	**9.8**
	0.05	5747	68.7	6.3
	0.01	4409	52.7	1.3
	0.005	3865	46.2	0.7
	0.001	2672	32.0	0.2

Table S16. The proportion of genes differentially expressed between species.

The false discovery rate (FDR) is estimated by 1,000 permutations. The permutation based *p*-value for the species effect is < 0.001 in all cases. We choose a *p*-value cutoff for each dataset/subset that ensures an FDR ~10%, shown in bold. The FDR for age-related genes is close to the FDR for all expressed genes (data not shown).

	***p*-value cutoff**	**Number of significant genes**	**% significant among all expressed**	**% FDR**
Dorsolateral PFC **39 humans *vs* 14 chimpanzees**	0.05	5506	69.2	38.4
	0.01	4480	56.3	19.4
	0.005	4109	51.6	14
	0.0025	**3790**	**47.6**	**9.8**
	0.001	3424	43.0	6
Dorsolateral PFC **14 humans *vs* 14 chimpanzees**	0.05	4384	55.1	16.7
	0.025	**3809**	**47.9**	**10.3**
	0.01	3185	40.0	5.2
	0.005	2772	34.8	3.1
	0.001	2049	25.7	0.9
Dorsolateral PFC **6 adult humans *vs* 5 adult chimpanzees**	**0.06**	**2743**	**34.5**	**10.3**
	0.05	2564	32.2	8.8
	0.01	1295	16.3	2.4
	0.005	983	12.4	1.4
	0.001	467	5.9	0.4
Dorsolateral PFC (2) [a] **6 adult humans *vs* 5 adult chimpanzees**	**0.055**	**2440**	**33.2**	**10.2**
	0.05	2366	32.2	9.3
	0.01	1195	16.2	2
	0.005	867	11.8	1.2
	0.001	400	5.4	0.2
S Frontal Gyrus **9 humans *vs* 9 chimpanzees**	0.05	4996	44.1	19.7
	0.015	**3332**	**29.4**	**10.4**
	0.01	2911	25.7	8.3
	0.005	2290	20.2	5.7
	0.001	1345	11.9	2.2
Brain regions **13 cortex *vs* 13 caudate nuc.**	0.05	8048	76.3	12.1
	0.04	**7866**	**74.6**	**10.2**
	0.01	6818	64.7	3.2
	0.005	6373	60.4	1.9
	0.001	5323	50.5	0.5
Mouse **9 *M.musculus* vs 9 *M.spretus***	0.05	4645	55.5	21.7
	0.01	**3251**	**38.9**	**8.9**
	0.005	2799	33.5	6.1
	0.001	1940	23.2	2.1

[a] Differential expression among adolescents and adults using the published the dataset from (1).

Table S17. Lineage assignment of human-chimpanzee differences.

The numbers of differentially expressed genes for which we can assign expression differences to the human lineage (human-specific) or chimpanzee lineage (chimpanzee-specific) using rhesus macaque as an outgroup (at one-sided Wilcoxon test $p<0.05$).

	Dataset	**Dorsolateral PFC**	**Dorsolateral PFC**	**S Frontal Gyrus**
	Number of humans vs chimpanzees compared	**14 *vs* 14**	**39 *vs* 14**	**9 *vs* 9**
Differentially expressed and age-related	Total	716	1058	977
	Human specific	287	398	309
	Chimpanzee specific	203	334	252
	% human specific	58.6	54.4	55.1
Differentially expressed and *not* age-related	Total	695	352	2355
	Human specific	235	115	657
	Chimpanzee specific	237	134	1028
	% human specific	49.8	46.2	39.0

Supporting Materials and Methods

Sample collection and hybridization to microarrays

Sample collection. All human *postmortem* brain tissue samples were obtained from the NICHD Brain and Tissue Bank for Developmental Disorders (NICHDBB) (Baltimore, MD, USA). 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 NICHDBB. No subjects with prolonged agonal state were used; cause of death is shown in Table S1. Chimpanzee samples were obtained from the Yerkes Primate Center (Atlanta, GA, USA), the Anthropological Institute & Museum of the University of Zürich-Irchel, (Zürich, Switzerland), and from the Biomedical Primate Research Centre (Rijswijk, Netherlands). 10 chimpanzee individuals were genotyped by sequencing the HVR1 region of the mitochondrial genome; all except one belonged to the Western chimpanzee population (Table S1; see Supporting Data at page 63). The dorsolateral PFC rhesus macaque samples were obtained from the German Primate Center (DPZ) (Goettingen, Germany), and for the superior frontal gyrus, from the SuZhou Experimental Animal Center (SuZhou, China). All non-human primates used in this study suffered sudden deaths for reasons other than their participation in this study and without any relation to the tissue used. For the mouse experiment, WSB/EiJ (*Mus musculus*) and SPRET/EiJ (*Mus spretus*) strain mice were purchased from the Jackson Laboratory (Bar Harbor, Maine, USA) and bred in our facility.

Tissue dissection. The human dorsolateral prefrontal cortex samples were dissected from the middle third of the middle frontal gyrus, from coronal slabs rostral to the head of the caudate nucleus, and from the equivalent region in the non-human primates. This is a cortical region approximately corresponding to Brodmann area 46. For all samples special care was taken to dissect primarily grey matter, but we cannot exclude the possibility of some contamination from white matter, and we assume the proportion of grey matter to be in the order of 90-95% of the total sample. The human caudate nucleus samples were taken from the central part of the nucleus in order to avoid any cross-contamination by adjacent brain structures. The superior frontal gyrus samples were taken from the cortical region approximately corresponding to Brodmann area 9 of the

prefrontal cortex; we aimed to make dissections that contain 2:1 grey:white matter (60-70% grey matter). For the mouse experiment we used whole brains. Expression profiles of adult mouse whole brains mainly reflect expression of the cerebral cortex compared to striatum or cerebellum (data not shown).

No samples showed any substantial RNA degradation, as measured using an Agilent Bioanalyzer (Agilent Technologies, Palo Alto, USA) indicating good tissue preservation. Detailed information for the human dorsolateral PFC samples, including age, sex, brain pH, *postmortem* interval, and death cause, are given in Table S1.

RNA extraction and chip hybridization. All experimental procedures have been previously described elsewhere (1). In brief, total RNA was extracted from approximately 100 mg of the dissected tissue sample (primates) or the whole brain (mice) using standard TriZOL® protocol with no modifications and purified with the QIAGEN® RNeasy MiniElute kit following the "RNA cleanup" protocol. RNA quality was assessed using the Agilent 2100 Bioanalyzer system. For each sample, 1 microgram (HG-U133P2 experiments), 2 micrograms (HuGene 1.0 ST experiments), or 5 micrograms (MG-430 2.0 experiments) of isolated total RNA was used as starting material for the standard Affymetrix eukaryotic target preparation protocol (see http://www.affymetrix.com/support/technical/byproduct.affx?product=hg-u133-plus, http://www.affymetrix.com/support/technical/byproduct.affx?product=moe430-20, http://www.affymetrix.com/products_services/arrays/specific/hugene_1_0_st.affx). Each sample was then hybridized to an Affymetrix® HG-U133 Plus 2.0 GeneChip array, an Affymetrix® Human Gene 1.0 ST array, or an Affymetrix® Mouse Genome 430 2.0 GeneChip array.

All raw expression data is deposited in the NCBI Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/; (4)) with the accession numbers GSE11512 (HG-U133P2 experiments), GSE11528 (MG-430 2.0 experiments) (accession number pending for the HuGene 1.0 ST experiments).

The effects of technical and clinical variables. To ensure that our analysis on the effects of age on expression is not confounded by technical or demographic variables, we first investigated the correlation between age and post-mortem interval (PMI), brain pH,

28S/18S ratio (a measure of RNA integrity), 5'-3' intensity slope (a measure of 5' degradation of transcripts), and sex, among the 39 human dorsolateral prefrontal cortex samples. We observe no significant correlation for any of these variables, with a single exception of Spearman correlation for *postmortem* interval (Table S13). However, this correlation is weak (*rho*=-0.32, p=0.05) and is not significant when corrected for multiple testing. Further, there is no significant relationship between PMI on RNA integrity (our results and previous results (5)). Thus, we do not expect variation in PMI among samples to have a significant influence on our results. Similarly in the superior frontal gyrus dataset, RNA quality levels among samples do not show any correlation with age (Spearman correlation test p>0.5), or difference among species (ANOVA F-test p=0.3).

We then explicitly estimated the effects of the RNA quality (28S/18S ratio) and brain pH and their possible interactions with age and species on each gene in the DLPFC human-chimpanzee dataset. We used multiple regression models and compared the results to 1,000 permutations of each factor (Table S12). Both factors affect the expression levels of ~25% expressed genes, and these proportions are significant in the permutation test (p<0.05). However, none of the interaction terms (age-pH, age-28s/18s, species-28s/18s) are significant (p>0.1). In addition, removing genes with significant pH or 28s/18s ratio effects from the dataset does not affect our principal results (Table S6).

Regarding cause of death, as mentioned above, all non-human subjects used in this study suffered sudden deaths for reasons other than their participation in this study and without any relation to the brain. For the human subjects, we have information for 37 out of the 39 individuals included in the dorsolateral PFC dataset (Table S1). Of the 37 individuals, 41% of cases are noted as asphyxia, drowning or asthma, 14% sudden infant death, 24% related to accidents, 14% of cases are related to cardiovascular disorders. In total, 34 out of 39 subjects have experienced fast deaths, without experiencing prolonged agonal state, which is known to affect both RNA integrity and brain pH (6). Importantly, there is no significant bias in distribution of rapid death cases with age (*rho*=0.24, p=0.12). This result agrees with our observations that tissue pH and RNA integrity do not show significant interaction with age (Table S12) and indicates that cause of death does not have a confounding effect on our analysis.

The effects of sex. To measure the possible effects of sex differences on brain gene expression, we used multiple regression models with age, sex and species (also see section *"The influence of age, species identity and sex on expression - a general view"*). In contrast to the highly significant age and species effects, neither the effect of sex, nor sex-age and sex-species interactions were significant (permutation test $p>0.3$) (Table S12). We see that sex affects only a small number of genes (13 genes at Bonferroni corrected $p<0.05$, in contrast to 2,522 genes showing age effects and 1,737 genes showing species effects). All 13 sex-related genes are located on the sex chromosomes. Comparable results were reported previously (7).

Thus, sex appears to have a very small effect on brain gene expression in this dataset. Nonetheless, to ensure that differences in sex between species do not affect our results, we based our main analysis in the dorsolateral prefrontal cortex dataset on a subset of 14 humans (Fig. S1C), such that these 14 human and 14 chimpanzee samples both have similar ages, and a comparable sex distribution across ages. In addition, we performed the analysis using only males, which yielded the same principal results (Table S6). Finally, we note that all humans in the superior frontal gyrus dataset are male, and results from this analysis are consistent with our other findings (Table 1). Taken together, these observations indicate that sex distribution differences among species do not affect our results.

Dataset preparation and the age-scale

All data preparation and analysis was conducted in the R statistical environment (8).

Chip definition files for the primate datasets. Chip definition (CDF) files contain information about the assignment of Affymetrix® expression analysis microarray probes to transcripts (see http://www.affymetrix.com/support/). For analyzing our expression data, we used a published "Custom CDF" file, an alternative to the standard CDF files provided by Affymetrix, that joins all probes from the HG-U133Plus2.0 microarray chip belonging to the same Ensembl gene into one probe set (http://brainarray.mbni.med.umich.edu/Brainarray/Database/CustomCDF/CDF_download.asp; (3); version 11). We will refer to all probe sets simply as "genes". The number of genes and probes in the CDF file, after excluding Affymetrix spike-in probes, are shown in Table S14.

For the analysis of chimpanzee and rhesus microarray data, we masked all probes that did not perfectly match the DNA sequences of these species (9). For this, we mapped sequences of all probes present on the HG-U133Plus2.0 or HuGene1.0 arrays to the human (hg18), chimpanzee (panTro2), and rhesus macaque (rheMac2) genomes using the BLAT alignment algorithm (http://genome.ucsc.edu/FAQ/FAQblat.html) (10). Based on these alignments, we constructed three new CDF files, described in Table S14 and available at http://www.picb.ac.cn/Comparative/data.html.

The "HG-U133P2 H-C Masked Custom CDF" is a subset of the original Ensembl Custom CDF, and contains probes matching both the human and chimpanzee genomes perfectly and at a single location (75% of the original array probes). The "HG-U133P2 H-C-R Masked Custom CDF" contains probes matching all three genomes perfectly and at a single location (31% of the original array probes). The "HuGene 1.0 H-C-R Masked Affymetrix CDF" file is a similar subset of the original Affymetrix CDF file (see http://www.affymetrix.com/support/) and contains 27% of the original array probes.

Chip definition files for the mouse dataset. As with the primate dataset preparation, we used the "Custom CDF" file that joins all probes on the MG-430_2.0 microarray chip that belong to the same mouse Ensembl genes into a single probe set ((3); version 11), and

removed Affymetrix spike-in probes. To compare *Mus musculus* and *Mus spretus* expression data we need a mouse masked CDF file that excludes probes with sequence differences between *M. musculus* (the species for which the microarray is designed) and *M. spretus*. Because the latter's genome is currently not available, we used a method which uses the discrepancy between the expression levels of probes within a probe set to identify probes that may contain sequence differences between the two genomes (Dannemann *et al*., in preparation; also see (11)). Briefly, a statistical test is run on all probes in a microarray dataset to estimate the degree of consistency of their expression levels with the other probes in the same probe set. A low consistency (a binding affinity difference) suggests a sequence difference for that probe. The test was run separately on three expression datasets containing 6 newborn, 6 infant or 6 adult *M.musculus* and *M.spretus*. Any probe with a significant binding affinity difference (at cutoff 0.2; described in Dannemann *et al*.) in any of the three datasets was discarded. 19% of the probes in the "MG-430_2.0 Ensembl Custom CDF" file were thus masked out (Table S14).

Preprocessing the human HG-U133P2 arrays. For the analysis of human brain regions, we extracted and summarized gene expression levels from the 39 Affymetrix .CEL files corresponding to the human dorsolateral prefrontal cortex samples and 13 to caudate nucleus samples using the "HG-U133P2 Custom CDF" file described above. For summarizing expression values we used the "rma" algorithm (R-Bioconductor "affy" package; http://bioconductor.org/packages/2.0/bioc/html/affy.html) which includes background correction, quantile normalization, log-2 transformation of expression levels, and summarizes expression levels of all probes in a probe set into one value using median-polishing (12). We decide on whether a gene is expressed above a background level on an array by testing the difference between signals from "perfect match" and "mismatch" probes from this array by using the Wilcoxon test ("mas5" method in the "affy" package) (12) and using $p<0.05$ as cutoff. We consider any gene with significant detection in >1/3 of individuals within either sample group (cortex or caudate nucleus) as expressed in that dataset.

Preprocessing the primate HG-U133P2 arrays. For the human-chimpanzee comparison, we extracted and normalized expression data from 39 human and 14 chimpanzee

dorsolateral PFC samples as described above, using the "HG-U133P2 H-C Masked Custom CDF" file (Table S14). We consider any gene with significant detection in >1/3 of human or chimpanzee individuals as expressed.

For the comparison among three primate species, we extracted and normalized expression data from 39 human, 14 chimpanzee and 9 rhesus macaque prefrontal cortex samples using the "HG-U133P2 H-C-R Masked Custom CDF" file. We consider any gene with significant detection in >1/3 of human or chimpanzee individuals as expressed. Here, we do not take macaque expression into account, because genes exclusively expressed in rhesus macaques would not be relevant to our main research focus: human-chimpanzee differences.

We also note that requiring significant detection in all individuals, or using an expression level based cutoff (genes with the highest 50% mean expression intensity), does not affect our findings (Table S6). We have also tested the possible effects of probe masking by a bootstrap: we randomly sampled 8 probes from each probe set with replacement, and repeated the full analysis with these datasets. We find that bootstrapping probes does not affect our findings (see legend in Table S6).

Preprocessing the primate HuGene1.0 arrays. Probe intensities were corrected for background using the "antigenomic" probes with the same GC content; the latter are used as an estimator of the unspecific background hybridization (http://www.affymetrix.com/support/technical/whitepapers/exon_background_correction_whitepaper.pdf). Probe intensities were then log-transformed and quantile normalized. Intensity values per transcript were calculated by median polishing. To determine whether the signal intensity of a given probe was above the expected level of background noise, we compared each probe's signal intensity to a distribution of signal intensities of the "antigenomic" probes with the same GC content (a GC-bin). For each GC-bin, except the ones with the most extreme GC content, the numbers of "antigenomic" probes are close to 1,000. We considered a probe signal as detected if its intensity is higher than 95% of the background probes' intensities (13). In each array, we considered a transcript as "detected" if more than 50% of probes and at least 8 probes per transcript were detected. We considered a transcript as "expressed" if it was detected in >1/3 of human or

chimpanzee individuals. 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 127 genes with multiple transcripts, we calculated the means across transcripts.

Preprocessing the mouse arrays. We extracted and normalized data from *M. musculus* and *M.spretus* arrays using the "MG430_2 M-S Masked Custom CDF" file (Table S14). Any gene with significant detection in >1/3 of individuals within either species was considered expressed.

The age scale, gestation period, and age-matched groups. To analyse age-related differences in gene expression among samples, we first decided on whether to use a linear or logarithmic age scale. Log transformed ages are frequently used in analysing developmental phenotypes with parametric models (*e.g.* (14-16)). This makes the relationship between age and phenotype more linear. This choice is related to an assumption of parametric regression models, that errors be normally distributed and homogeneous among samples (see Chapters 13.6 and 13.7 in (17)). Sokal and Rohlf suggest log transformation as a primary option for continuous variables where this assumption does not hold. In all our datasets, when the linear age scale is used, error sizes are larger at the early age; this reflects the very rapid change in expression levels during early brain development (Fig. 1*D*). This bias can be considerably reduced when log-transformed age is used: In the human-chimpanzee dataset, Pearson correlation between age and squared errors of cubic expression-age regression models decreases from -0.26 to -0.12 with the use of log age. Hence, unless indicated otherwise, we used the logarithm-transformed age scale throughout the analysis. Nevertheless, we find similar results if the linear age scale is used (Table S6).

Another issue in assigning ages to individuals is whether the ages should be counted starting from birth or from conception. It is likely that the postnatal gene expression changes represent continuation of the ontogenetic processes initiated in the prenatal period. To account for this, we calculated individuals' age starting from inferred time of conception (for an example see (14)). The known average gestation periods are 280, 229 and 165 days in humans, chimpanzees and rhesus macaques, respectively (18). However, gestation times for humans and chimpanzees may be calculated in different ways, and the

real human-chimp difference may be around 20 days (as pointed out by an anonymous reviewer). Further, actual gestation times may vary among individuals. As using a smaller gestation time difference than the actual one is conservative with respect to identifying human-neoteny, we conducted our main analyses using 280 days as gestation time for both humans and chimpanzees. Further, we conducted additional analysis using 20 and 51 days as gestation time difference between the species, and obtained the same results (Table 1).

Finally, in the dorsolateral PFC dataset, our analysis on species differences between humans and chimpanzees could be biased by unequal sampling (39 vs. 14 individuals) or by the differences in age distribution between the species. We therefore performed our main analyses using a subset of 14 human samples matching the ages of the chimpanzee samples as close as possible (Fig. S1*C*). Although this approach limits our power, *e.g.* in identifying age effects (Table S15), we obtain the same results as with the full human dataset (Table 1). Unless stated otherwise, our analyses of this dataset are based on the 14 humans. We adopt a similar approach for the cortex-caudate nucleus comparison and choose 13 cortex samples age-matched with the caudate nucleus samples (Fig. S1*D*).

The influence of age, species identity and sex on expression - a general view

We used a number of methods to describe the influence of age, species identity and sex on gene expression levels in our datasets.

Principle components analysis. We scaled the variance of each gene to one, and used the "prcomp" function (R "stats" package) to calculate the principle components from the human-chimpanzee DLPFC dataset. The first component shows the strongest correlation with age, while the second component is most strongly correlated with species differences (Figure 1A).

Multiple regression models of age, sex and species. To obtain an overall view of the effects of different factors on gene expression, for each gene, we fit a multiple regression model with age, species identity and sex, as well as all possible interactions among these. We calculated the proportion of expressed genes showing significant main factor or interaction effects at F-test $p<0.05$. We then compared the results to 1,000 permutations of either of the 3 factors (Table S12). We find significant age and species effects, as well as an age-species interaction; but we see no significant sex-related effects on expression. Note that a previous study on humans and mice also did not find large expression differences between the brains of males and females (7).

Proportion of variance explained by age, sex and species. In addition to assessing the number of genes that are significantly affected by different factors, we can also measure how much of total expression variation can be accounted by different factors. For this, for each gene, we fit separate regression models for age, species identity or sex in the human and the chimpanzee dataset. For species and sex, we applied linear regression models, whereas for age, we used third degree polynomial models. We also estimated the proportion of variance explained by the age and species identity factors together, with separate age-related parameters for humans and chimpanzees.

The proportion of expression variance attributable to factor *f* for gene *i* was calculated as:

$$R^2_{i,f} = 1 - \mathrm{RSS}_{i,f} / \mathrm{RSS}_{i,null}$$

where $\mathrm{RSS}_{i,f}$ is the residual sum of squares from the regression model with factor *f*, and $\mathrm{RSS}_{i,null}$ is the residual sum of squares from the null model, in other words, the variance

multiplied by n – 1, where n is the sample size. We further estimated a significance level for the mean proportion of variance explained across all 7,958 genes expressed in the prefrontal cortex, R^2_f, by randomly permuting the factor (*f*) assignments across samples 1,000 times to obtain a distribution of randomized mean proportions: R^{2*}_f. The permutation *p*-value was defined as the frequency of permutations in which $R^{2*}_f >= R^2_f$. The FDR was defined as the ratio between the median of the 1,000 R^{2*}_f values and R^2_f.

Age and species explain expression variance at significantly higher levels than in random permutations in all datasets (Table S2). Surprisingly, sex explained variance at levels *significantly lower*. We hypothesized the following: The sex factor in our dataset is correlated with neither age nor species (Pearson r = 0.02 and -0.04, respectively); however, when sex is randomly permuted, the new arrangements *f** can happen to be more strongly correlated with either age or species (half the time |r*| > 0.09 and 0.07, respectively). Given the large influence of age and species on expression variation, *f** can frequently explain more variance than sex itself. This effect can be shown by simulation (data not shown). After removing the influence of age and species on expression by using residuals from a multiple regression model, we find the variance explained by sex is neither higher nor lower than random permutations (Table S2; Fig. 1*B*).

The global expression measure. To obtain a one-dimensional summary measure of expression variation among a set of samples across all expressed genes, we first calculated a matrix of Euclidean distances among the samples. Then we collapsed these distances into a single dimension by means of monotonic regression, a type of multidimensional scaling ("isoMDS" algorithm with k=1, R "MASS" package (19)). We call this the global expression measure. Subjects with similar values of the global expression measure can be inferred to resemble each other in terms of their expression patterns across all genes, compared to those with less similar values. The scale of this measure is arbitrary, and it represents only the most dominant source of variation among samples. We used this measure in Figure 1*D*, where we calculated each subjects' global expression value as a percent of the range between the maximum and minimum values among all subjects, separately for humans and chimpanzees.

Tests for age, species and lineage effects - gene-by-gene analysis

The R code used to apply the tests described below is available at http://www.picb.ac.cn/Comparative/data.html.

Testing the influence of age – model selection. We first determine the effect of age on expression by choosing, for each gene, the best polynomial regression with age as predictor and expression level as response. For this, we used families of polynomial regression models and the "adjusted r^2" criterion (20). Namely, we fit a third degree regression model with age to each gene:

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

where Y_{ij} is the expression level for gene i with i = 1,..., m and subject j with j = 1,..., n, A_j is the age of the subject j, and ε_{ij} is the error term.

We further calculate the six possible submodels of equation (1), *e.g.* $Y_{ij} = \beta_{0i} + \beta_{1i} A_j + \varepsilon_{ij}$, $Y_{ij} = \beta_{0i} + \beta_{1i} A_j + \beta_{2i} A_j^2 + \varepsilon_{ij}$, *etc*.

Then we compare all seven models to the null model:

$$Y_{ij} = \beta_{0i} + \varepsilon_{ji} \quad (2)$$

by means of an F-test. We choose the model with the highest "adjusted r^2" value as the best choice (20). If the model is significant in the F-test at a predetermined significance cutoff, we consider this gene age-related ("age+").

We make three notes at this point:

First, the highest adjusted r^2 reflects the amount of variance explained by the model, but penalized by the number of parameters of the model – this approach therefore avoids overfitting (choosing models with unnecessary parameters).

Second, we do not test each model against each other, but use the adjusted r^2 criterion to compare models. This method is both time-efficient and effective. We show in Supporting Note 1 that if a model is significant and has the highest adjusted r^2, this is good indication that it would also be at least marginally better (at F-test p<0.10) if directly compared to alternative models, 96% of the time.

Third, we do not attempt to correct for multiple testing with respect to the number of models tested, because the tests with different models conducted on a single gene are not independent. Furthermore, we use a permutation test to assess overall significance of the result (see below).

We also tried the cubic smoothing spline as an alternative to third degree polynomials in order to describe the relationship between expression and age ("smooth.spline" function, R "stats" package). We tested seven smoothing spline models with degrees of freedom ranging from 2 to 8, following the procedure described above for the polynomial regression models. We chose the best model (the best number of degrees of freedom) by comparing each model to the null model [equation (2) above] using the F-test, and choosing the one with the highest adjusted r^2. In the dorsolateral prefrontal cortex dataset, at the same significance cutoff (p<0.045) we find 73% age-related genes using spline models, compared to 71% when using polynomial regression models. Furthermore, 99% (5,578 of 5,653) of genes identified as age-related using the regression models are similarly identified using the spline models.

Testing the influence of age – FDR. We used permutations to estimate the false discovery rate and decide on the F-test significance cutoff for each dataset. We randomly permuted age assignments across individuals within a species 1,000 times, repeated the regression analysis for all genes and recorded the F-test p-values. For each p-value cutoff (*e.g.* 0.05) we calculated the number of genes with p-values below the cutoff in the 1,000 permutations. The median of this distribution is our false positive estimate. The ratio between this estimate and the original number of significant genes is the false discovery rate (FDR). Note that this FDR estimation is conservative (21). In addition to FDR, the frequency of permutations in which the number of significant genes is equal to or greater than that observed can be considered a p-value for a transcriptome-wide age-effect. Table S15 shows the results of this age-test in a number of datasets. For each dataset, we chose a p-value cutoff that would fix the FDR close to 10%. Table S8 shows the overlap between age-related genes identified in different datasets.

Testing species differences – model selection. For each gene, we determined differential expression between two species by multiple regression: comparing the outcomes of regression models with and without a species term (an approach similar to the one described in (22)). For non-age-related genes, this approach is equivalent to a t-test. For age-related genes, we have already chosen a regression model with age for human, and we determine whether incorporating chimpanzee-specific parameters into this regression model significantly improves it. Specifically, using the F-test, we compare the null model

(*i.e.* the regression model with age but no chimpanzee-specific parameter) with a family of models containing a chimpanzee specific intercept. In addition, for each age-related parameter of the null model, we included a chimpanzee-specific term. For example, if the null model (best fitting model with age for human) for gene *i* was a linear one:

$Y_{ij} = \beta_{0i} + \beta_{1i}A_j + \varepsilon_{ij}$,

we compare it to the two alternative models:

$Y_{ij} = \beta_{0iC} + \beta_{1i}A_j + \varepsilon_{ij}$, and

$Y_{ij} = \beta_{0iC} + \beta_{1iC}A_j + \varepsilon_{ij}$.

where β_{0iC} and β_{1iC} are the chimpanzee-specific intercept and slope. We choose the model with the highest adjusted r^2. We then decide on the significance of the chosen model using an FDR-based *p*-value cutoff.

Testing species differences – FDR. We measured the false discovery rate using the same approach as the age-test, by running 1,000 random permutations of species identity assignments across samples. However, in order to preserve the structure in the data with respect to age as much as possible, and only randomize species identity, we permuted humans and chimpanzees of similar age. Specifically, in a permutation, each time a chimpanzee is assigned to the first group, the human with the closest age to that chimpanzee is assigned to the second group (see Fig. S1*C*). Consequently, the age distributions of the two randomly formed groups are similar to the original age distributions. We then apply the multiple regression test to these randomly chosen groups and compare the results to the original one. We accept the median of the 1,000 permutations' number of significant genes as the random expectation (Table S16). The *p*-value for a transcriptome-wide species effect is the frequency of permutations that yield as many significant genes as observed. The transcriptome-wide *p*-values and FDR's for age-related and non-age-related genes are similar (data not shown).

Table S16 shows the results of this species-test on a number of datasets. For each dataset, we chose a *p*-value that would fix the FDR close to 10%. Table S8 shows the overlap between differentially expressed genes identified in different datasets. Note that using the same sample size (6 adult humans and 5 adult chimpanzees) and the same statistical

criteria to define differently expressed genes, the proportion of differentially expressed genes in our DLPFC dataset is very similar to a previously published dataset (1) (Table S16). Further, the majority of differently expressed genes overlap between the two datasets (50-52%) (see Table S8).

Human or chimpanzee lineage assignment. We used the rhesus macaque as an outgroup to assign expression differences between humans and chimpanzees to the human or to the chimpanzee evolutionary lineages. We note that this test was limited by the relatively small number of rhesus samples in the dorsolateral PFC dataset (n=9), the limited age-range of rhesus macaques, and especially the loss of power for expression level estimation: more than two-thirds of probes are masked due to sequence mismatches among species (Table S14).

For each gene, we determined whether the rhesus expression profiles are closer to the human or to the chimpanzee profile using the following procedure: First, we estimate the expression-age curves for the human and chimpanzee groups, separately. The regression model for the curve is chosen based on the age-test, described above. For non-age-related genes, the curves are simply the species means. Second, we calculate the absolute distances of each rhesus individual's expression level to each curve. Third, we test if the distances to the human curve are larger than the chimpanzee curve by a one-sided paired Wilcoxon test (Table S17). Note that because we do not have a prior hypothesis, the real significance level is twice as low (observed p=0.05 corresponds to actual p=0.10).

We calculated FDR for the lineage test using the permutation method described for the species-test: we permute species identities while conserving the age-distributions of the groups. The FDR for both the dorsolateral PFC and the superior frontal gyrus datasets was between 80%-95%, although the test was significant (more genes assigned to lineages than in random permutations, p<0.05 in both datasets). This is surprising, given that there is considerable overlap between the two brain region datasets (DLPFC and SFG) for human- specific and chimpanzee-specific genes identified in these datasets, better than for the species-test results (Table S8). We therefore do not try to control FDR in this test but use different p-value cutoffs for assigning genes to lineages (Table 1).

To test robustness of the evolutionary lineage assignment we conducted an alternative, although not independent, lineage assignment procedure: Using the test for expression differences between species described above, we classified genes as human-specific if they showed significant differences between humans and chimpanzees and between humans and rhesus macaques, but not between chimpanzees and rhesus macaques (and *vice versa* for chimpanzee-specific genes). Among 351 lineage-specific genes identified by this method (at two-sided $p<0.05$ for both tests), 198 were also identified by the first method, and all but two were classified consistently. Further, we find the same excess of human-neotenic genes using this lineage assignment method (Table S6).

The heterochrony test for gene expression

The R code used to apply the heterochrony test is available at http://www.picb.ac.cn/Comparative/data.html.

The principle of the heterochrony test. In this test, we are comparing the ontogenetic expression patterns of one taxon with another taxon (*e.g.* the human *versus* the chimpanzee) with respect to timing differences. We assume that both taxa follow an ontogenetic trajectory of the same general shape. For a gene, whenever there is a difference in ontogenetic rates, or timing, between the two taxa (*e.g.* humans maturing twice as slow as chimpanzees), the expression values of one taxon will correspond to expression values of younger or older stages of the other taxon. Therefore, transforming the ages of the second taxon (*e.g.* multiplying the chimpanzees' ages by a certain factor) can produce an improved fit between the two groups' expression distributions with age (Fig. S3). Any gene for which this improvement is statistically significant, we term heterochronic. Further, depending upon the direction of the age transformation, we classify heterochronic genes as neotenic or accelerated in one taxon *versus* another. For example, genes exhibiting delayed maturation in human *versus* chimpanzee are called neotenic in human. About the use of the term neoteny with reference to morphometric literature, see Supporting Discussion.

A critical assumption of the heterochrony test is that trajectories of expression change with age are comparable in shape across species. Morphometric comparisons of human and ape skull development has found largely different growth trajectories between these species (23). However, our results show extensive positive correlation between ontogenetic expression trajectories – not only between humans and chimpanzee, but also between primates and rodents. This indicates that the assumption of similar developmental trajectories between humans and chimpanzees is applicable to the vast majority of genes expressed in brain (Table S4 and Fig. S2).

Defining models to compare human and chimpanzee. We start by determining, for each gene, the best cubic smoothing spline curve that describes expression changes with age in the reference species. We use human as the reference species in our analysis, but using chimpanzee yields similar results (Table S6). We use spline models here rather than cubic

polynomial regression, because spline curves are more flexible and thus can fit the expression data more closely, which facilitates comparing the human curve and chimpanzee expression values. However, using polynomial regression also does not affect the results (data not shown).

Given the human spline curve, we try alternative models on each gene (transforming the chimpanzee age scale or shifting the chimpanzee mean expression level) to improve the fit between chimpanzee expression values (Y_{ij}) and the human curve (Y'_{ij}). For example, if the best fitting model for human for gene *i* was linear, the estimate of the expression value for individual *j* for gene *i* would be:

$$Y'_{ij} = \beta_{0i} + \beta_{1i}A_j,$$

where the parameters β_{0i} and β_{1i} describe the human curve, and A_j is the log2-transformed age of the chimpanzee *j*. The sum of squared differences between these estimates (Y'_{ij}) and the real chimpanzee expression values (Y_{ij}) represents, among different sources of variation, differences in developmental rates of the two species:

$$\sum_j (Y_{ij} - Y'_{ij})^2 = \sum_j (Y_{ij} - [\beta_{0i} + \beta_{1i}A_j])^2. \quad (3)$$

(a) Model A - the age-shift model: Now, instead of using the real ages for the chimpanzees, we redefine these. Assuming that the shape of curves are similar in the two species, we seek a constant (C) to add to all chimpanzee subjects' ages that will compensate for the difference between the rates of ontogenetic change:

$$A_{Cj} = C + A_j.$$

Note that C, or the age-shift, is also on the log2-transformed age-scale; hence adding C to A_j is equivalent to multiplying non-log-transformed ages by 2^C. C is therefore a ratio between developmental rates. $C > 0$ implies that chimpanzee ontogenesis is accelerated with respect to human -- this is synonymous to human ontogenesis being delayed, or neotenic. The reverse is true if $C < 0$.

In order to find a good estimate of C given the expression data, we use an iterative non-linear least squares algorithm, "NL2SOL" ("nls" function, R "stats" package) and search for a value of C (staring from 0) that minimizes the sum of squared differences between

the observed chimpanzee expression values and expression values estimated from the human curve at the transformed chimpanzee ages (A_{Cj}):

$\sum_j (Y_{ij} - Y'_{ij})^2 = Y_{ij} - [\beta_{0i} + \beta_{1i}(C + A_j)])^2$.

In other words, we estimate the age-shift, C, that minimizes the differences between rates of age-related expression change in humans and chimpanzees. The sign of C will imply whether human ontogenesis is delayed or accelerated relative to chimpanzee.

(b) Model B - the expression-shift model: For each age-related gene, we can estimate an age-shift, C, as described above. But it is also conceivable that differences between the chimpanzee expression values (Y_{ij}) and the estimated expression values from the human model (Y'_{ij}), may be better explained by a shift of average expression levels between the two species (M), that is fixed across ages. Following the above example, if the human expression-age model for gene *i* were linear, we would estimate the M that minimizes:

$\sum_j (Y_{ij} - Y'_{ij})^2 = \sum_j (Y_{ij} - [\beta_{0i} + \beta_{1i}A_j + M])^2$.

which is simply the mean difference between Y_{ij} and Y'_{ij}.

(c) Model AB - the combined model: This is the combination of models A and B. We again use the "NL2SOL" algorithm to jointly estimate the age-shift (C) and expression-shift (M), by finding the values that minimize:

$\sum_j (Y_{ij} - Y'_{ij})^2 = \sum_j (Y_{ij} - [\beta_{0i} + \beta_{1i}(C + A_j) + M])^2$.

The search for both parameters starts from 0. Occasionally the NL2SOL algorithm cannot converge on C and M estimates in the AB model. This happens when the C and M have equivalent effects on $Y_{ij} - Y'_{ij}$, so that the algorithm cannot disentangle the two.

Fig. S3 presents an example of how an age-shift can improve the fit between the human spline curve and the chimpanzee expression values.

Choosing significant models. After the age-shift (C) and expression-shift (M) are estimated by the Models A, B, and AB, we use the F-test to compare each of these models to the null model (equation 3 is an example for the null model). We also compare the models A and B to model AB by the F-test. A significant F-test result indicates that

the tested model is both more parsimonious and explains significantly greater amount of variation than the null model.

Having estimated C and M using these three different models and having estimated their significance using the F-test, we proceed to classify genes into neotenic or accelerated categories -- genes showing patterns compatible with human neoteny or acceleration. Specifically, for a gene to be classified as heterochronic, first, the F-tests should provide support for the significance of the age-shift, and second, this support should at least as be as strong as that for the expression-shift -- there is no evidence for heterochrony if the expression-shift (Model B) is significantly better than Models A and AB. We defined the following set of conditions ("condition set 1") to decide on whether there is support for heterochrony, and which model's C estimate to use -- A or AB:

a) Model AB is the best among the three models (has the highest adjusted r^2) and is significantly better than the null model ($p<0.05$, F-test). For these genes, the C estimate from model AB is the best estimate for the age-shift.

b) Model A is the best among the three models and is significant ($p<0.05$, F-test). We consider the C estimate from model A as our best estimate.

c) Model B is the best among the three models and significant ($p<0.05$, F-test). But model A is also significant ($p<0.05$), and model AB is not better than model A ($p>0.05$), which implies that the expression-shift is not significantly better than the age-shift (following the J-test logic described in (24)). We then accept the C estimate from model A.

d) Model B is the best model and significant ($p<0.05$, F-test). But model A is also significant ($p<0.05$). In addition, model AB does not converge (see above) suggesting that models A and B are alike. In this case, model A is considered as good as model B, and we accept the C estimate from model A.

Once we decide on which C estimate to use, we classify a gene as neotenic or accelerated in human *versus* chimpanzee, depending on whether C is positive or negative, respectively. We further use the F-test *p*-value as a measure of the significance of this

estimate. In the main analysis we use $p<0.05$ as cutoff (unless otherwise indicated). We discuss the false negative and positive rates of the test below.

Classifying genes in phylo-ontogenetic categories, false positive and negative rates

Significance criteria for classification. We assign genes into four phylo-ontogenetic categories based on the type of the heterochrony found between humans and chimpanzees (neoteny or acceleration), and on the placement of the corresponding expression change to an evolutionary lineage (the human or the chimpanzee lineage) (Fig. 3*A*, Table 1). Based on these parameters, human-neotenic genes are the ones that show a neotenic shift in human *versus* chimpanzee, and the expression changes are assigned to the human evolutionary lineage using the rhesus macaque. Human-accelerated genes show an accelerated pattern in human *versus* chimpanzee, and expression changes are assigned to the human lineage. Chimpanzee-accelerated genes show a neotenic pattern in human *vs* chimpanzee, but expression changes are assigned to the chimpanzee lineage. Therefore, we infer that for such genes the chimpanzee expression changes with age are accelerated with respect to human. Chimpanzee-neotenic genes are the ones that show an accelerated pattern in human *versus* chimpanzee and the expression changes are assigned to the chimpanzee lineage (Table 1 and Table S6).

We used three sets of significance cutoffs to define phylo-ontogenetic categories. *(a)* The default was using FDR=10% for the age and species differences, one-sided binomial test $p<0.05$ for the lineage assignment test, and F-test $p<0.05$ for the heterochrony test; we are unable to directly control FDR for the latter two tests. *(b)* $p<0.10$ for all four tests. *(c)* $p<0.01$ for all four tests (Table 1). $p<0.10$ may appear too relaxed for a cutoff, but as we may have high false negative rates in our tests, using relaxed cutoffs may increase our power to observe species-specific heterochrony.

Testing the bias towards human-neoteny. We conducted two tests to assess whether the number of human-neotenic genes is greater than expected by chance, given the results of the heterochrony and the lineage tests (Table 1 and Table S6): *(a)* among all genes assigned to the human lineage (human-specific genes), we asked if the number of human-neotenic genes is larger than the number of human-accelerated genes, *(b)* among all genes classified as neotenic, we asked if human-specific genes are more than chimpanzee-specific genes. In both cases we used one-sided binomial tests.

Excess of human-neotenic genes under alternative criteria or assumptions. In line with the human neoteny hypothesis, we found a significant excess of human-neotenic genes in both tests, which cause human expression profiles to resemble younger chimpanzees (Fig. S4). Furthermore, the two independent datasets from different brain regions (DLPFC and SFG) showed the same results. To test the robustness of this result, we further conducted parallel analyses with different sets assumptions and criteria with respect to the number of individuals used per species, the gestation time, sex of individuals used in analysis, choice of the lineage assignment test, choice of logarithmic or linear age scales, and the choice of reference species in the heterochrony test (Table S1, Table S6). We also performed the analyses by excluding genes affected by technical factors, changing the criteria for defining which genes are expressed, and randomly selecting probes within probe sets (Table S6).

Finally, we applied two alternative conditions sets to define heterochrony besides the one used above. *(a)* We defined a more stringent set of conditions, requiring support for the age-shift be unequivocal compared to the support for an expression-shift ("condition set 2"; described in Supporting Note 2). *(b)* We only used the age-shift (C) estimate from Model A, ignoring Models B and AB ("condition set 3"). Regardless of the method, we find the same excess of neotenic genes (Table S6).

In these analyses, we focused on the dorsolateral PFC dataset, where we have considerably larger sample sizes. Nonetheless, the same tests applied to the SFG dataset result in the same excess of human-neotenic genes (data not shown).

False positive and false negative rates of phylo-ontogenetic assignment. We used several approaches to estimate the false positive and the false negative rates of the phylo-ontogenetic classification in the human-chimpanzee dorsolateral PFC dataset (Table S7).

First, we estimated the false positive rate by comparing humans against humans. For this we simply compared 14 humans to 14 other humans with a similar age distribution, using either group as reference species. Here, although we do not expect any, we find that 25-30 genes showing significant results with respect to the species-difference, heterochrony, and lineage tests using the default significance cutoffs (Table S7). This can be contrasted with the 299 genes in the real human-chimpanzee comparison with the same sample size.

These results suggest that our overall result have a false positive rate in order of 10% at these significance cutoffs (Table S7). The false positive rate increases to ~30% at relaxed *p*-value cutoffs ($p<0.1$ for all tests). Also, the test appears to be biased towards identifying acceleration, rather than neoteny (this is also apparent in the simulation discussed below).

Second, we examined the false negative rate by simulating neoteny and acceleration. We again compared 14 humans with another group of 14 humans, but after dividing the ages of the first group by 2. We can identify only 27% of age-related genes as delayed, that is, significantly "group-one-human"-neotenic or "group-two-human"-accelerated (Table S7), while 100% is expected without noise and any bias in our tests. As expected, the number of genes showing significant heterochrony in direction opposite to the simulated one is negligible. Given the simulation conditions and assuming a two-fold difference in the developmental rate affecting the entire transcriptome, these results suggest high false negative rate of our tests: in the order of 75%. When we simulate acceleration in the reference species, our tests identify a larger number of accelerated genes, 33%. This suggests that, in addition to the high false negative rates, our tests may be biased towards finding more accelerated than neotenic genes.

We also applied the same procedure by simulating a 1.5 times rate difference (instead of 2) between group-one and group-two. This time we estimate 20% of age-related genes as delayed when simulating neoteny, and 26% as accelerated when simulating acceleration.

The important point is that if a neotenic shift of such magnitudes (1.5-2 times faster development in chimpanzees relative to humans) were affecting the entire transcriptome, the power of our study would be sufficient to identify a much greater proportion of the neotenic genes than discovered in the real data (Table S7). Therefore, a developmental change of this magnitude that equally affects all genes is improbable.

Comparing mice with primates and human caudate nucleus with cortex

Age and species effects in the mouse. We conducted analyses on the influence of age and species identity in the *Mus musculus* and *Mus spretus* (musculus-spretus) dataset following the same procedure as for the human-chimpanzee dataset. First, using the 8,362 expressed genes in the musculus-spretus dataset, we calculated the proportion of variance explained by age and species using third degree polynomial models and linear models, as described above, and conducted 1,000 permutations. We also applied the test for age and species effects, as described above, conducted 1,000 permutations for each test, and assigned genes to classes at FDR=10% (Table S15 and Table S16).

Note that the proportion of expression variance explained by age in mice is substantially larger than both the primate dorsolateral prefrontal cortex and superior frontal gyrus experiments (Table S2). This can potentially be explained by the larger number of postnatal developmental changes in the mouse brain (14), the controlled environment of the mice reducing the impact of environmental factors on the individual expression variation, as well as the low genetic variation within the mouse groups.

Conservation of age-related changes in mice and primates. Using mouse gene expression changes with age, we sought to assess the level of conservation of age-related expression changes in rodents and primates. To do so, we first obtained a list of orthologous human and mouse genes from the Ensembl Biomart (http://www.ensembl.org/biomart/; (25)). Among the 22,207 orthologs in the dataset, we chose the "one-to-one" orthologs, which left us with 14,941 human-mouse orthologs. We compared the primate and mouse experiment results using three different approaches. First we compared the overlaps between age-related genes in the two experiments using the Fisher's exact test. This comparison yields a small but significant overlap between age-related genes identified in primates and mice (Table S8).

Second, we directly compared age-related changes in the different datasets. For this, (*i*) we grouped individuals in the human-chimpanzee dataset into three age groups that would approximate the developmental states of the three mouse age groups: Newborns (younger than 1 year of age), adolescents (between 10 and 14 years of age), and adults

(between 25 and 50 years of age). 7, 3, and 2 human prefrontal cortex samples, and 6, 3, and 2 chimpanzee prefrontal cortex samples fell in these three age groups, respectively. (*ii*) For each age group in each species, for every expressed gene, we calculated the mean expression level among individuals in an age group. (*iii*) Using these average gene expression values in the three age groups, we compared all taxa (humans, chimpanzees, *M. musculus* and *M. spretus*) in pairs. Namely, for each pair of species, for a set of expressed orthologs, we calculated the Pearson correlation between the mean expression levels of the three age groups. Table S4 shows the median correlation coefficient for each pair of taxa among 2,599 age-related (age+) orthologs and 388 non-age-related (age-) orthologs (genes which are age-related or non-age-related in both primates and rodents) (Table S4). For age- genes, we find either weak or no correlation between primate and rodent expression profiles. In contrast, the average correlation between humans and the mouse species for age+ genes is comparable to that between humans and chimpanzees.

Third, for Fig. S2, we first normalized the expression levels of the 2,599 human-mouse age+ orthologs in each species' dataset to mean 0 and variance 1. We then clustered these genes into four groups based on the expression levels of the 3 human age groups, using K-means clustering ("kmeans" function in the R "stats" package). Finally we plotted the mean expression levels of genes in all four clusters for all four species.

Caudate nucleus ontogenesis compared with prefrontal cortex. We processed the dataset containing expression profiles of the 39 human dorsolateral prefrontal cortex and 13 human caudate nucleus samples following the same steps as for the human and chimpanzee comparison. First, among the 10,543 expressed Ensembl genes, we identified 7,437 age-related genes (at F-test $p<0.045$, FDR=10%) using the 39 prefrontal cortex samples. Next, we tested each gene for expression level differences between the prefrontal cortex and caudate nucleus using the species-test (Table S16). For each gene showing significant age and species effects, we tested the existence of a developmental shift in the caudate nucleus expression-age profile compared to prefrontal cortex using the heterochrony test. We applied the tests on a subset of 13 prefrontal cortex samples age-matched to the 13 caudate nucleus samples (Fig. S1*D*), as well as all 39 cortex samples. Details are presented in Table S5.

Characterizing human-neotenic genes

Gene Ontology analysis. To test whether human-neotenic genes are overrepresented in particular biological functions compared to genes in the other three phylo-ontogenetic categories, we used functional annotation provided by the Gene Ontology (GO) Consortium (26) and a statistical algorithm developed for testing gene distributions along the GO taxonomies, func_hyper (http://func.eva.mpg.de; (27)). We used Ensembl Biomart annotation (http://www.ensembl.org/biomart/; (25)) to assign genes to GO categories. The Gene Ontology is structured as three taxonomies: biological process, molecular function and cellular component, each of them containing large numbers of nested functional groups. The func_hyper program runs hypergeometric tests for enrichment across all GO groups within a taxonomy; for each GO group, it tests whether there is a larger overlap between genes in that GO group and a set of genes of interest, compared to a control set of genes. Next, given the number of enriched GO groups, func_hyper runs a permutation test and calculates an overall *p*-value for the whole taxonomy, indicating whether the patterns of enrichment within a taxonomy are out of ordinary (27). This ensures that the enrichment signal is reliable given multiple hypothesis testing. The three taxonomies are not independent, and here we only used the "biological process" taxonomy.

We compared human-neotenic genes to all genes assigned to the other three phylo-ontogenetic categories. We find no global enrichment, however, we do find a number of enriched GO groups identified both in the dorsolateral PFC and superior frontal gyrus datasets (Table S10 and Table S11). The lack of global significance in GO analysis with these gene-selection criteria is most likely due to relatively small numbers of genes identified as human-neotenic. Indeed, when using relaxed gene detection criteria with no cutoff for the number of expressed probes in a probe set, with the resulting approximately three-fold increase in numbers of analyzed genes, we again find the same GO groups as before, and the global enrichment of human-neotenic genes among GO categories is marginally significant ($p<0.10$). Similarly, when we applied the same approach to test functional specificity of 5,653 age-related genes compared to all 7,958 expressed genes in the human-chimpanzee dorsolateral prefrontal cortex dataset (Table S3), we find highly significant overall enrichment ($p=0.003$).

Grey vs. white matter specificity. We determined whether human-neotenic genes are enriched among genes highly expressed in grey matter or the white brain matter, compared to genes in the other three phylo-ontogenetic categories. For this, we used data from a study where expression profiles of grey matter or white matter had been measured in two regions of the human prefrontal cortex, Brodmann areas 9 and 47 (2). Using this data, we defined sets of genes with grey or white matter specific expression by the following procedure: We downloaded the table titled "588_probesets.xls" in (2), and for each of the 22,215 Affymetrix® HG-U133A array probe sets in the table, we *(i)* calculated the mean percentage of detected samples in the two regions, *(ii)* calculated the mean log ratio of white matter to grey matter in the two regions, *(iii)* selected probe sets which had a mean detection frequency larger than 50% and an absolute log ratio larger than 3 (in the original study the authors had used 1.5; we chose to be more stringent), *(iv)* mapped Affymetrix probe sets to human Ensembl genes using the Ensembl Biomart (http://www.ensembl.org/biomart/), *(v)* removed genes that occurred in both sets. The resulting 1,155 and 578 genes were used as grey matter and white matter-specific gene sets, respectively, among all 9,892 expressed genes in this dataset. Using the Fisher's exact test, we compared human-neotenic genes with genes in the three other phylo-genetic categories, with respect to their overlap with grey- or white-matter specific gene sets. As background, we used all expressed genes in the Erraji-Benchekroun *et al.* dataset.

Expression divergence across life. We determined how human-chimpanzee expression divergence changes with age using the following procedure: For each gene, *(i)* we calculated the regression model of expression on age for the 14 human age-matched subjects, and separately for the chimpanzees; *(ii)* calculated the absolute expression level difference between the two curves at intervals of 0.1 age points on the log-transformed age scale. *(iii)* We normalized these differences to mean 0 and standard deviation 1 per gene. *(iv)* For each gene set of interest (*e.g.* human-neotenic genes), we bootstrapped over genes 10,000 times. We thus calculated the median and 95% confidence interval for divergence at each point across the age-scale (Fig. 3*C*).

Supporting notes and discussion

Supporting Note 1: Using the adjusted r2 criterion in model selection

In the section titled *"Testing the influence of age - model selection"* above, we argue that, among possible polynomial regression models of expression change with age, the model which has highest adjusted r^2, and thus the smallest F-test *p*-value compared to the null model should be relatively parsimonious. This implies that such a model is also significantly better than other possible models with age, if directly compared to these alternative models. A simple simulation based on the below R code shows this the case: If a regression model containing factors "predictor1" and "predictor2" is significant at F-test $p<0.05$ and has a lower F-test *p*-value than a model with only "predictor1" (when both are compared to the null model), then the "predictor1"+"predictor2" model is also better than the "predictor1" model (100% of the time at F-test $p<0.15$; 96% of the time at F-test $p<0.10$; %84 of the time at F-test $p<0.05$ in 500 runs).

```
PERM = 500
Mat = matrix(,PERM, 3)
i = 1
while (i < (PERM+1)) {
  dependent = 1:40
  predictor1 = rnorm(40)
  predictor2 = rnorm(40)
  p1 = anova(lm(dependent ~ 1), lm(dependent ~ predictor1))$P[2]
  p12 = anova(lm(dependent ~ 1), lm(dependent ~ predictor1 + predictor2))$P[2]
  if (p12 < 0.05 & p12 < p1) {
    p3 = anova(lm(dependent ~ predictor1), lm(dependent ~ predictor1 + predictor2))$P[2]
    Mat[i, ] = c(p1, p12, p3)
    i = i + 1
} }
1-sum(Mat[,3]>0.1)/nrow(Mat)
```

Supporting Note 2: Alternative criteria for defining neoteny

If for a gene, the F-test results point to significantly stronger support for an age-shift than for an expression-shift, we consider such a gene neotenic or accelerated in the narrow sense. The difference from the default set of criteria described in section "*Choosing significant models*" (condition set 1), is that we now require support for the age-shift be unequivocal compared to the support for an expression-shift. We define three possible conditions to classify genes as such. A gene fulfilling one of these conditions is considered as showing heterochrony (see Table S6).

***a*)** Model AB (the combined age-shift and expression-shift model) is significantly better than the null model ($p<0.05$, F-test), and is the best among the three models (has the highest adjusted r^2). Model AB is also better than model B (the expression-shift model) when the two are directly compared ($p<0.05$). This means that the age-shift (C) in model AB is significant. We then consider the C estimate from model AB as significant and as our best estimate for an age-correction. If the C estimate is > 0, we classify the gene as neotenic, if < 0, as accelerated.

b) Model A (the age-shift model) is significantly better than the null model ($p<0.05$, F-test) and the best among all three models. Model AB is also significantly better than model B ($p<0.05$) but not A ($p>0.05$). Together, these results indicate that the expression-shift is trivial, whereas the age-shift (C) results in a good fit between the chimpanzee expression values and the human curve.

c) Model A is best compared to the null model ($p<0.05$, F-test). Meanwhile model AB does not converge (see above) suggesting that models A and B are alike. In this case, model A is good enough only when it is significant ($p<0.05$) while model B is not ($p>0.05$). For genes fulfilling these conditions, we use the C estimate from model A.

Supporting Discussion: The heterochrony test and different modes of heterochrony

Following Gould's 1977 book (28), Alberch *et al.* defined three general modes of heterochrony: *(a)* neoteny *versus* acceleration, referring to different rates of

morphological change, *(b)* post-displacement *vs* pre-displacement, referring to differences in the onset of morphological change, and *(c)* progenesis *vs* hypermorphosis, referring to differences in the duration of morphological change (29). Neoteny, post-displacement, and progenesis are alternative paths leading to paedomorphosis, whereas acceleration, pre-displacement, and hypermorphosis lead to peramorphosis.

Although these definitions have since been discussed and even changed (*e.g.* (30), reviewed in (31)), the model in general is widely accepted as a framework to analyze developmental changes (31-33). It is therefore appealing to know how our expression heterochrony test performs given these different modes of heterochrony. This is especially interesting because an increase in growth rates and extension of growth periods, such as seen in human brain or body growth relative to the chimpanzee, have been argued to represent human hypermorphosis, rather than human neoteny (32, 34, 35).

We ran a simulation to observe how the heterochrony test assesses these different modes. We randomly generated data from Gaussian distributions, simulating a heterochronic change: in each case, the reference species exhibited either neoteny, post-displacement, or hypermorphosis (following Alberch *et al.*'s nomenclature) relative to the second species. Examples of simulated data are shown in Fig. S5 below. We then applied the heterochrony test to these datasets.

In simulations of both neoteny and post-displacement, the heterochrony test estimates a positive age-shift in the age-shift model (model A), as expected. Therefore, using "condition set 1" (the default), we would classify both patterns as "neotenic."The question here is whether we could have distinguished neotenic and post-displacement patterns, for instance, by modifying our test to detect differences in the shapes and slopes of trajectories. This is in principle possible, but the sample sizes in our datasets would not be sufficient to confidently make this distinction.

A more complicated result appears from simulations of hypermorphosis. We found that the age-shift model (model A) estimates a negative age-shift for examples of expression hypermorphy in our simulation. However, the combined age-shift and expression-shift model (model AB) estimated a positive age-shift, and explained a larger proportion of the variance. Therefore, using "condition set 1", we would classify hypermorphic gene

expression patterns as shown in Fig. S5 as "neotenic". This is in fact expected, as our algorithm was designed to classify all patterns in which human expression level is delayed relative to the chimpanzee as "neotenic".

However, in contrast to neoteny and post-displacement (following Alberch *et al.*'s nomenclature), which are types of paedomorposis, hypermorphosis is a type of peramorphosis, or overgrowth (33). So does the large number of human-neotenic genes identified in our human-chimpanzee comparison, in reality reflect hypermorphic expression patterns? If this were the case, we would expect no excess of human-neotenic genes to be found when genes are classified using only results from model A ("condition set 3"), ignoring the expression-shift. In contrast, Table S6 shows that using "condition set 3" has no large effect on the excess of human-neotenic genes.

To further understand the influence of possible hypermorphic patterns on our results, we built hierarchical clustering trees based on human and chimpanzee expression data, using all age-related and differentially expressed genes, and using human-neotenic genes. In the gene expression tree based on all age-related and differentially expressed genes, newborn humans and chimpanzees cluster together, and the two species' adults also cluster (Fig. S4*A*). If hypermorphic patterns have a large influence among our list of human-neotenic genes, human adult brain gene expression would be mainly "overgrown" for these genes. Adult humans should then to form a clade distinct from all other groups in the human-neotenic gene expression tree. Instead, we find here that adult humans cluster with juvenile chimpanzees, to the exclusion of adult chimpanzees and macaques (Fig. S4*B*). This is compatible with paedomorphosis in human expression levels, for human-neotenic genes. The same pattern is found in Fig. 69 in Gould (28), based on morphometric data.

Most phenotypic characters are likely to be affected by combinations of different heterochronic patterns (29, 31). Thus, for most characters, heterochrony analysis will reveal only average patters. This is also true for gene expression differences: The patterns we detect at the single gene level are average tendencies. It is nevertheless interesting that this average tendency is biased towards delayed development, and results in human paedomorphosis. For historical reasons, we have chosen to use such the term "neoteny" for such a tendency. However, future studies using larger sample sizes, and improved

knowledge of regulatory processes, may allow distinguishing different modes of expression heterochrony that characterize human ontogenesis.

Supporting Data: Chimpanzee HVR1 sequences

>Neonat1_0_days

GGGCGAGGATGGATTTGACTGTAATGTGCTATGTACGGTGTGTGATTGTATGTACTATGTTTTGTCAAGG
GGGGGTAGGTTTGTTGGTATCGGGGTGGGGGAGAAGTGTCGTTGGAGTTGTGTTTTATGTTTGATAGTTA
GAGGTTGATTGTTGTGCATGCTTGTAAGCATGGGGTGAAGGTTTTAATGTGGGGTGGGTTTTATGTGTTAT
AGGTAGTTGGGTAATTATGGTACCGTACAATATTCATGGTGGCTGGCAGTAATGTACGAAATACATAGCG
GTTGTAATGAATGAGCCAGTACTTAGGTGGTACTTAAATTTGCTTCCCCATGAAAGAA

>Neonat2_1_days

GGGACGAGGATGGATTTGACTGTAATGTGCTATGTGCGATGTACGGTTGTATGTACTATGTACTGTTAAA
AAGGTATAGGTTTGTTGGTATCGGGGTGGGAGAGGGGTGTCTTTGGAGTTGCATTTTATGTGTGACAGTT
GGGGGTTGATTGTTGTGCGTGCTTGTAAGCATAGGGTAGAGGTTGTTATGTGGGGGTGGGTTTAATGTAT
TATAGGTAGTTGAGTGATTATAGTACTGTACAATATTCATGGTGGCTGGCAGTAATGTACGAAATACATA
GCGGTTGTAATGAATAGGCCAATACTTAGGTGGTACTTAAATTTGCTTCCCCATGAAAGAA

>Guno_8_days

GGGCGAGGATGGATTTGACTGTAATGTGCTATGTACGGTGTGTGATTGTATGTACTATGTTTTGTCAAGG
GGGGGTAGGTTTGTTGGTATCGGGGTGGGGGAGAAGTGTCGTTGGAGTTGTGTTTTATGTTTGATAGTTA
GAGGTTGATTGTTGTGCGTGCTTGTAAGCATGGGGTGAAGGTTTTAATGTGGGGTGGGTTTTATGTGTTAT
AGGTAGTTGGGTAATTATGGTACCGTACAATATTCATGGTGGCTGGCAGTAATGTACGAAATACATAGCG
GTTGTAATGAATGAGCCAGTACTTAGGTGGTACTTAAATTTGCTTCCCCATGAAAGAA

>MowgliII_40_days

GGGCGAGGATGGATTTGACTGTAATGTGCTATGTACGGTGTGTGATTGTATGTACTATGTTTTGTCAAGG
GGGGGTAGGTTTGTTGGTATCGGGGTGGGGGAGAAGTGTCGTTGGAGTTGTGTTTTATGTTTGATAGTTA
GAGGTTGATTGTTGTGCATGCTTGTAAGCATGGGGTGAAGGTTTTAATGTGGGGTGGGTTTTATGTGTTAT
AGGTAGTTGGGTAATTATGGTACCGTACAATATTCATGGTGGCTGGCAGTAATGTACGAAATACATAGCG
GTTGTAATGAATGAGCCAGTACTTAGGTGGTACTTAAATTTGCTTCCCCATGAAAGAA

>Gantzi_45_days

GGGCGAGGATGGATTTGACTGTAATGTGCTATGTACGGTGTGTGATTGTATGTACTATGTTTTGTCAAGG
GGGGGTAGGTTTGTTGGTATCGGGGTGGGGGAGAAGTGTCGTTGGAGTTGTGTTTTATGTTTGATAGTTA
GAGGTTGATTGTTGTGCGTGCTTGTAAGCATGGGGTGAAGGTTTTAATGTGGGGTGGGTTTTATGTGTTAT
AGGTAGTTGGGTAATTATGGTACCGTACAATATTCATGGTGGCTGGCAGTAATGTACGAAATACATAGCG
GTTGTAATGAATGAGCCAGTACTTAGGTGGTACTTAAATTTGCTTCCCCATGAAAGAA

>Mirante_186_days

GGGCGAGGATGGATTTGACTGTAATGTGCTATGTACGGTGTGTGATTGTATGTACTATGTTTTGTCAAGG
GGGGGTAGGTTTGTTGGTATCGGGGTGGGGGAGAAGTGTCGTTGGAGTTGTGTTTTATGTTTGATAGTTA
GAGGTTGATTGTTGTGCATGCTTGTAAGCATGGGGTGAAGGTTTTAATGTGGGGTGGGTTTTATGTGTTAT
AGGTAGTTGGGTAATTATGGTACCGTACAATATTCATGGTGGCTGGCAGTAATGTACGAAATACATAGCG
GTTGTAATGAATGAGCCAGTACTTAGGTGGTACTTAAATTTGCTTCCCCATGAAAGAA

>Herman_4361_days

GGGCGAGGAGGGGTTTGACTGTAATGTGCTATGTACGGTGTATGGTTGTATGTACTATGTTCTGTCAAGG
GGAGATAGGTCTGTTGGTATCGGGGTGGGGGAGGGGTGTCGTTGGAATTGTGTTTTATGTTCGACAGTTG
GGAGTTGATTGTTGTGCGTGCTTGTAAGCATGGAGTGAAGGTTTTGATGTGGGAGTGGATTTTATGTACT
ATAGGTAGTTGGGTGATTATGGTACTGTACGATATTCATGGTGGCTGGCAGTAATGTACGAAATACATAG
CGGTTGTAATGAATGAGCCAGTACTTAGGTGGTACTTAAATTTGCTTCCCCATGAAAGAA

>Koos_4415_days

GGGCGAGGATGGATTTGACTGTAATGTGCTATGTACGGTGTGTGATTGTATGTACTATGTTTTGTCAAGG
GGGGGTAGGTTTGTTGGTATCGGGGTGGGGGAGAAGTGTCGTTGGAGTTGTGTTTTATGTTTGATAGTTA
GAGGTTGATTGTTGTGCGTGCTTGTAAGCATGGGGTGAAGGTTTTAATGTGGGGTGGGTTTTATGTGTTAT
AGGTAGTTGGGTAATTATGGTACCGTACAATATTCATGGTGGCTGGCAGTAATGTACGAAATACATAGCG
GTTGTAATGAATGAGCCAGTACTTAGGTGGTACTTAAATTTGCTTCCCCATGAAAGAA

>Japie_4480_days

GGGCGAGGATGGATTTGACTGTAATGTGCTATGTACGGTGTGTGATTGTATGTACTATGTTTTGTCAAGG
GGGGGTAGGTTTGTTGGTATCGGGGTGGGGGAGAAGTGTCGTTGGAGTTGTGTTTTATGTTTGATAGTTA
GAGGTTGATTGTTGTGCGTGCTTGTAAGCATGGGGTGAAGGTTTTAATGTGGGGTGGGTTTTATGTGTTAT
AGGTAGTTGGGTAATTATGGTACCGTACAATATTCATGGTGGCTGGCAGTAATGTACGAAATACATAGCG
GTTGTAATGAATGAGCCAGTACTTAGGTGGTACTTAAATTTGCTTCCCCATGAAAGAA

>Reba_16131_days

GGGCAAGGATGGATTTGACTGTAATGTGCTATGTACGGTGTGTGATTGTATGTATTATGTTTTGTCAAGGG
AGGGTAGGTTTGTTGGTATTAGGGTGGGGGAGAAGTGTCGTTGGAGTTGTGTTTTATGTTCGATAGTTGG
GGGTTGATTGTTGTGCGTGCTTGTAAGCATGGGGTGAAGGTTTTAATGTGAGGTGGATTTTATGTGCTATA
GGTGGTTGGGTGATTATGGTACTGTACAATATTCATGGTGGCTGGCAGTAATGTACGAAATACATAGCGG
TTGTAATGAATGAGCCAGTACTTAGGTGGTACTTAAATTTGCTTCCCCATGAAAGAA

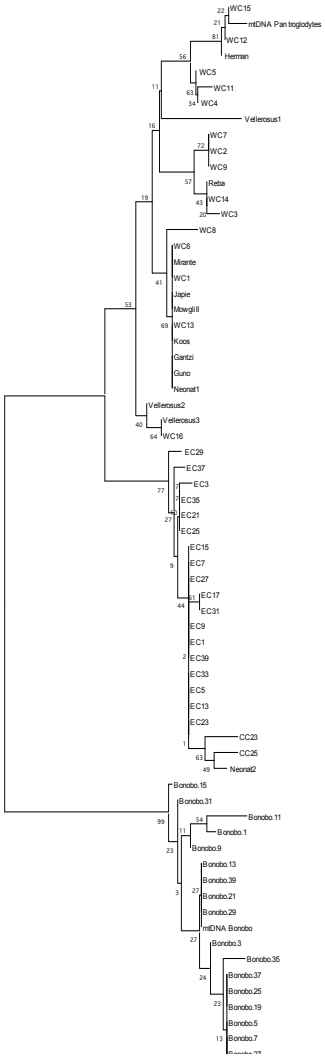

Figure S6: Chimpanzee phylogenetic tree based on HVR1

The mitochondrial HVR1 locus was successfully sequenced in 10 chimpanzee individuals. A neighbour joining tree was calculated with the gamma distribution (gamme = 0.32) using these 10 sequences (labelled by each individual's name in the figure), as well as HVR1 sequences from chimpanzees of Western (WC), Eastern (EC) and Central (CC) populations, and bonobos (Anne Fischer, unpublished data). Note that only one individual (Neotenat2, 1 day old) groups together with Central and Eastern chimpanzees; the other nine cluster with Western individuals.

Supporting References

1. Khaitovich, P., Hellmann, I., Enard, W., Nowick, K., Leinweber, M., Franz, H., Weiss, G., Lachmann, M., & Paabo, S. (2005) Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* 309: 1850-1854.
2. Erraji-Benchekroun, L., Underwood, M. D., Arango, V., Galfalvy, H., Pavlidis, P., Smyrniotopoulos, P., Mann, J. J., & Sibille, E. (2005) Molecular aging in human prefrontal cortex is selective and continuous throughout adult life. *Biol Psychiatry* 57: 549.
3. Dai, M., Wang, P., Boyd, A. D., Kostov, G., Athey, B., Jones, E. G., Bunney, W. E., Myers, R. M., Speed, T. P., Akil, H., *et al.* (2005) Evolving gene/transcript definitions significantly alter the interpretation of genechip data. *Nucl Acids Res* 33: e175-.
4. Edgar, R., Domrachev, M., & Lash, A. E. (2002) Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucl Acids Res* 30: 207-210.
5. Ervin, J. F., Heinzen, E. L., Cronin, K. D., Goldstein, D., Szymanski, M. H., Burke, J. R., Welsh-Bohmer, K. A., & Hulette, C. M. (2007) Postmortem delay has minimal effect on brain rna integrity. *J Neuropathol Exp Neurol* 66: 1093-1099.
6. Tomita, H., Vawter, M. P., Walsh, D. M., Evans, S. J., Choudary, P. V., Li, J., Overman, K. M., Atz, M. E., Myers, R. M., & Jones, E. G. (2004) Effect of agonal and postmortem factors on gene expression profile: Quality control in microarray analyses of postmortem human brain. *Biol Psychiatry* 55: 346.
7. Rinn, J. L., Rozowsky, J. S., Laurenzi, I. J., Petersen, P. H., Zou, K., Zhong, W., Gerstein, M., & Snyder, M. (2004) Major molecular differences between mammalian sexes are involved in drug metabolism and renal function. *Dev Cell* 6: 791-800.
8. Gentleman, R., Carey, V., Bates, D., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., *et al.* (2004) Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol* 5: R80.
9. Khaitovich, P., Enard, W., Lachmann, M., & Paabo, S. (2006) Evolution of primate gene expression. *Nat Rev Genet* 7: 693-702.
10. Kent, W. J. (2002) Blat---the blast-like alignment tool. *Genome Res* 12: 656-664.
11. Khaitovich, P., Muetzel, B., She, X., Lachmann, M., Hellmann, I., Dietzsch, J., Steigele, S., Do, H., Weiss, G., Enard, W., *et al.* (2004) Regional patterns of gene expression in human and chimpanzee brains. *Genome Res* 14: 1462-1473.
12. Gautier, L., Cope, L., Bolstad, B. M., & Irizarry, R. A. (2004) Affy--analysis of affymetrix genechip data at the probe level. *Bioinformatics* 20: 307-315.
13. Clark, T., Schweitzer, A., Chen, T., Staples, M., Lu, G., Wang, H., Williams, A., & Blume, J. (2007) Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biology* 8: R64.
14. Clancy, B., Darlington, R. B., & Finlay, B. L. (2001) Translating developmental time across mammalian species. *Neuroscience* 105: 7-17.
15. Lu, L., Airey, D. C., & Williams, R. W. (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.

16. Shupe, J. M., Kristan, D. M., Austad, S. N., & Stenkamp, D. L. (2006) The eye of the laboratory mouse remains anatomically adapted for natural conditions. *Brain Behav Evol* 67: 39-52.
17. Sokal, R. R. & Rohlf, F. J. (1995) *Biometry* (Freeman, New York).
18. de Magalhães, J. P. (2006) *Anage Database*, build 9. http://genomics.senescence.info/species/index.html.
19. Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*.
20. Faraway, J. (2002) *Practical regression and ANOVA using R*. http://www.stat.lsa.umich.edu/~faraway/book/
21. Xie, Y., Pan, W., & Khodursky, A. B. (2005) A note on using permutation-based false discovery rate estimates to compare different analysis methods for microarray data. *Bioinformatics* 21: 4280-4288.
22. Storey, J. D., Xiao, W., Leek, J. T., Tompkins, R. G., & Davis, R. W. (2005) Significance analysis of time course microarray experiments. *Proc Natl Acad Sci USA* 102: 12837-12842.
23. Mitteroecker, P., Gunz, P., Bernhard, M., Schaefer, K., & Bookstein, F. L. (2004) Comparison of cranial ontogenetic trajectories among great apes and humans. *J Human Evol* 46: 679-698.
24. Davidson, R. & MacKinnon, J. G. (1981) Several tests for model specification in the presence of alternative hypotheses. *Econometrica* 49: 781-793.
25. Hubbard, T. J. P., Aken, B. L., 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.
26. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T.*, et al.* (2000) Gene Ontology: Tool for the unification of biology. *Nat Genet* 25: 25-29.
27. Prüfer, K., Muetzel, B., Do, H., Weiss, G., Khaitovich, P. R., Pääbo, S., Lachmann, M., & Enard, W. (2007) Func: A package for detecting significant associations between gene sets and ontological annotations *BMC Bioinformatics* 8: 41.
28. Gould, S. J. (1977) *Ontogeny and Phylogeny*.
29. Alberch, P., Gould, S. J., Oster, G. F., & Wake, D. B. (1979) Size and shape in ontogeny and phylogeny. *Paleobiology* 5: 296-317.
30. Rice, S. (1997) The analysis of ontogenetic trajectories: When a change in size or shape is not heterochrony. *Proc Natl Acad Sci USA* 94: 907-912.
31. Klingenberg, C. P. (1998) Heterochrony and allometry: The analysis of evolutionary change in ontogeny. *Biol Reviews* 73: 79-123.
32. Shea, B. T. (1989) Heterochrony in human evolution: The case for neoteny reconsidered. *Am J Phys Anthropol* 32: 69-101.
33. McKinney, M. L. & McNamara, K. J. (1991) *Heterochrony: The Evolution of Ontogeny*.
34. McNamara, K. J. (1997) *Shapes of Time*.
35. Rice, S. H., Minugh-Purvis, N., & McNamara, K. (2002) in *Human Evolution Through Developmental Change*.