

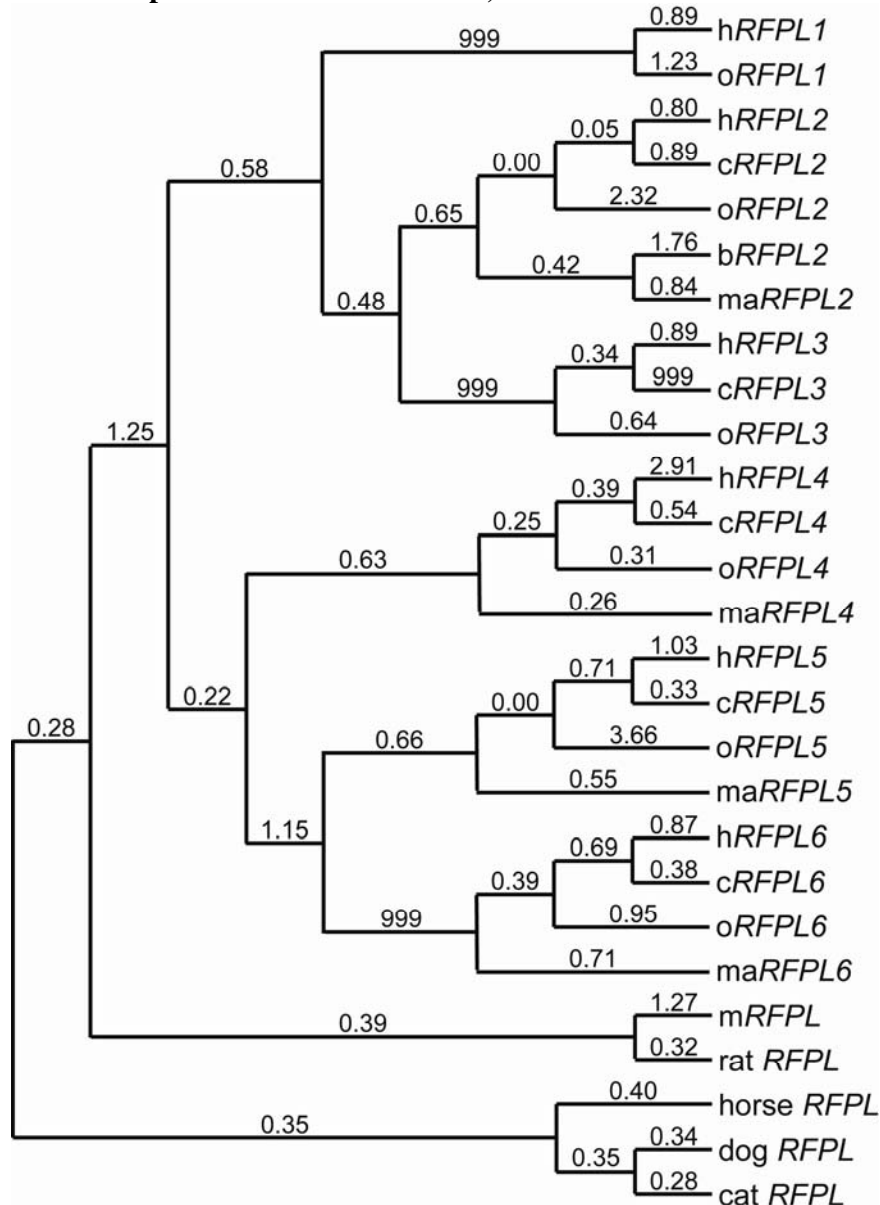
Supplemental Data

Evolutionary Forces Shape the Human *RFPL1,2,3* Genes

toward a Role in Neocortex Development

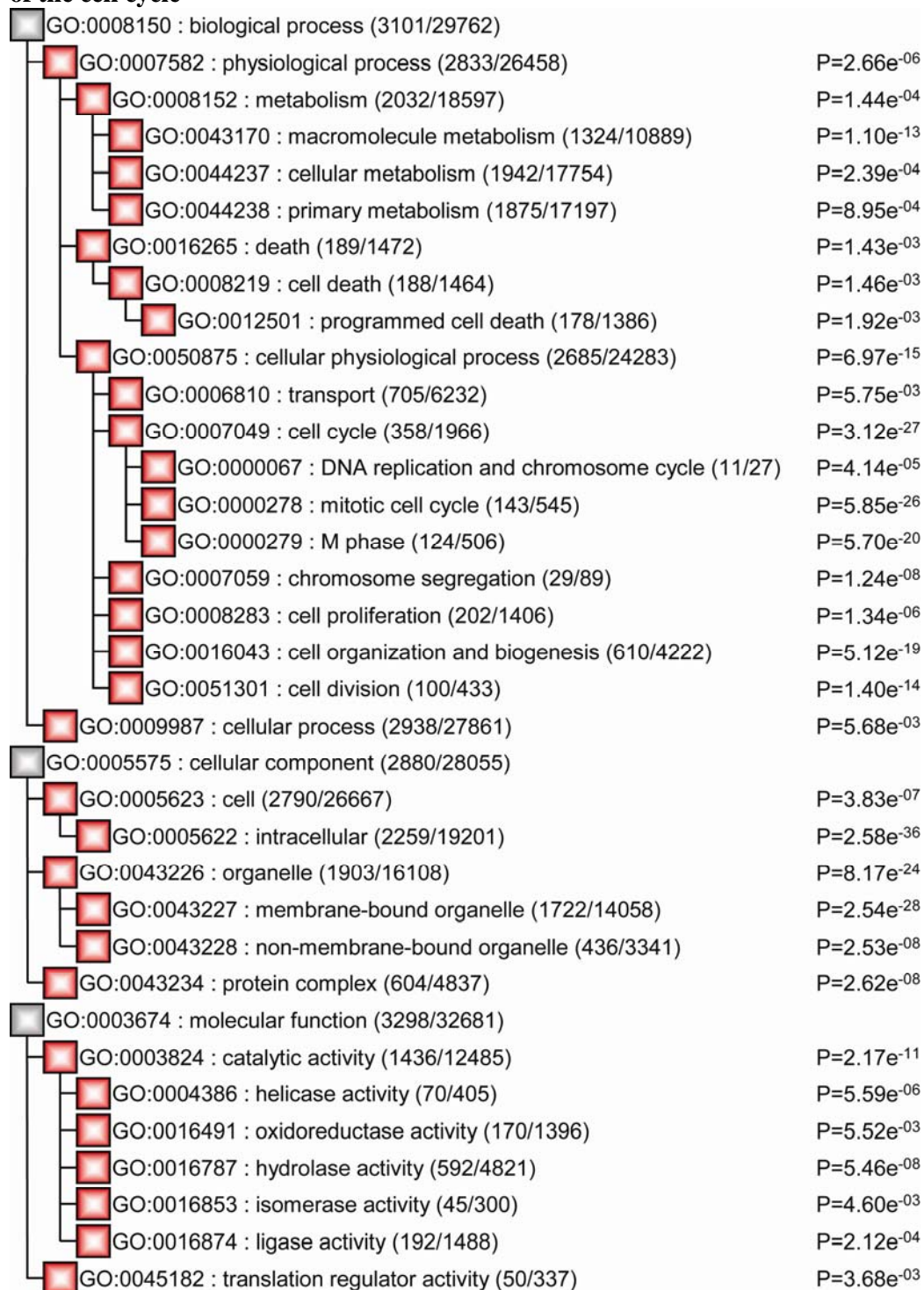
Jérôme Bonnefont, Sergey I. Nikolaev, Anselme L. Perrier, Song Guo, Laetitia Cartier, Silvia Sorce, T r se Laforge, Laetitia Aubry, Philipp Khaitovich, Marc Peschanski, Stylianos E. Antonarakis, and Karl-Heinz Krause

Figure S1. Phylogenetic tree based on RFPL coding sequences encoding the likely functional proteins in Laurasiatheria, Glires and Catarrhini



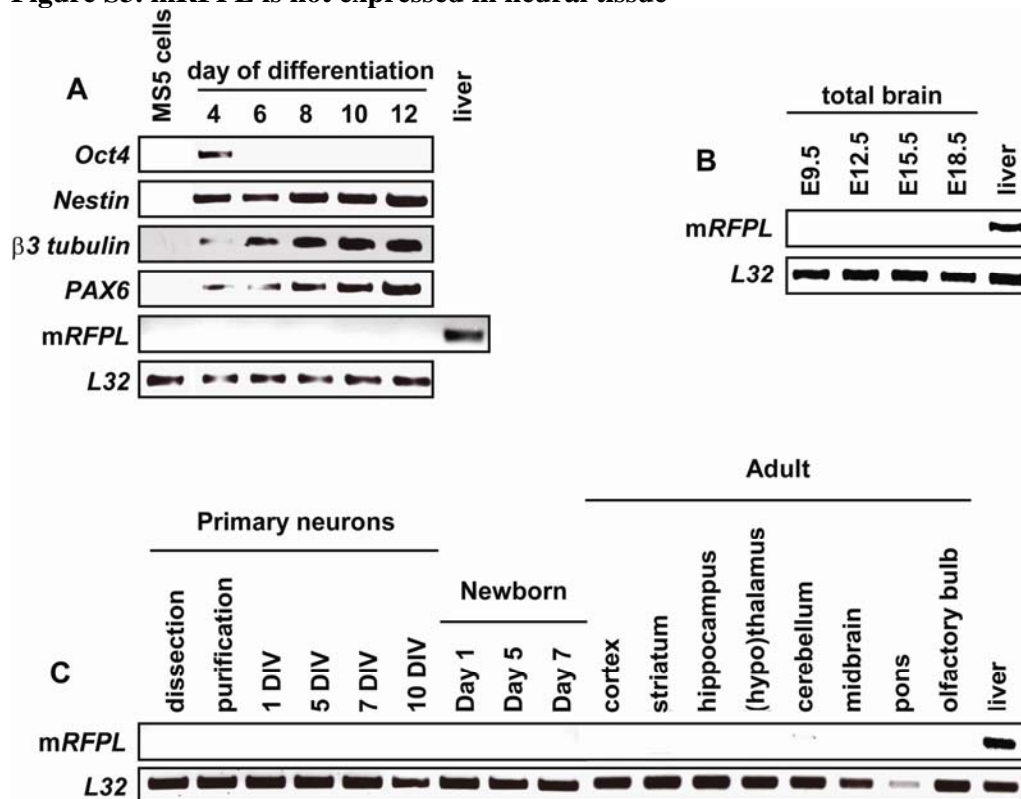
Maximum K_A/K_S values for each branch are indicated.

Figure S2. Pax6 highly significantly modifies gene clusters implicated in the regulation of the cell cycle



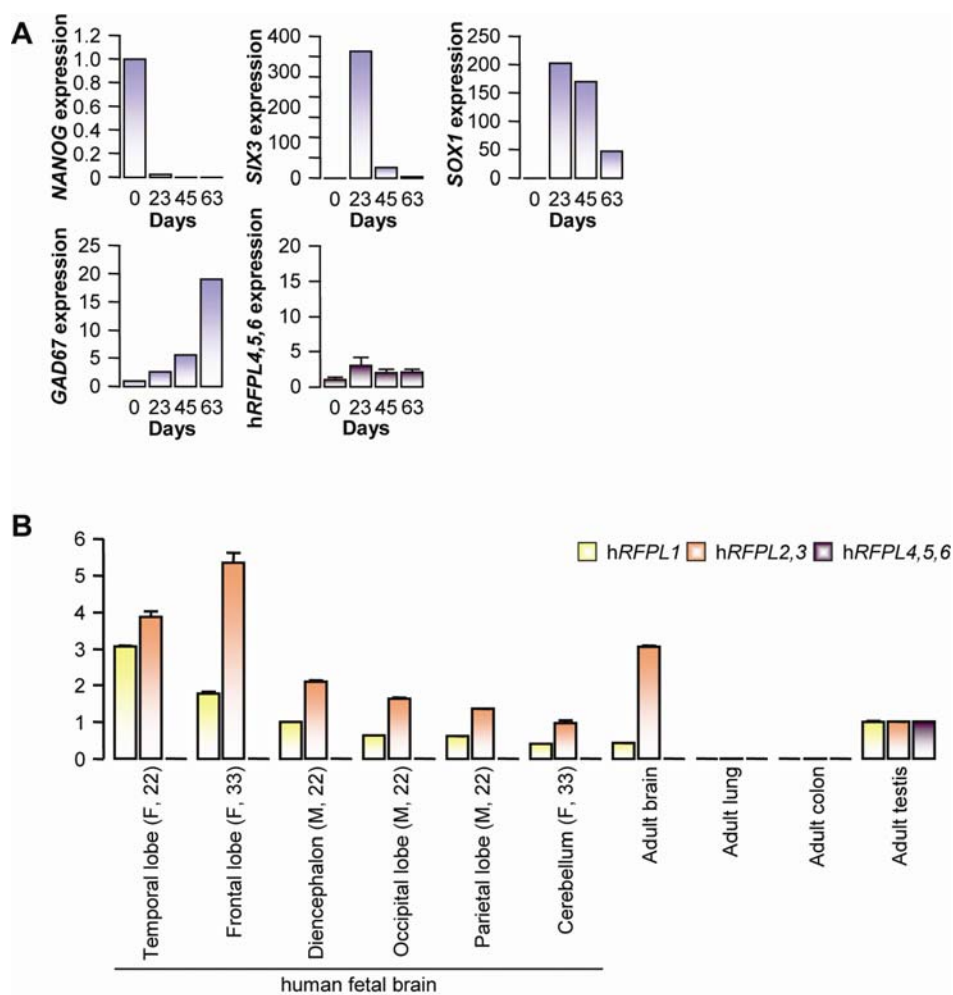
Pax6-elicited changes in gene cluster expressions observed in the biological process taxonomy defined by the Gene Ontology Consortium were determined using the GeneSpring GX 7 software. In parentheses are indicated the number of genes of the cluster that showed an altered expression over the total number of genes of the cluster.

Figure S3. mRFPL is not expressed in neural tissue



mRFPL expression was assessed by semi-quantitative end-point RT-PCR (A) during neurogenesis in differentiating murine D3 embryonic stem cells, (B) during neurogenesis in fetal brain, (C) during *in vitro* maturation of primary cortical neurons, and in newborn and adult mouse brain. Liver was used as a PCR positive control.

Figure S4. hRFPL4,5,6 genes are not expressed during neurogenesis *in vitro* and *in vivo*



Transcript levels were assessed by real-time PCR and normalized to the cDNA level in adult testis. (A) hRFPL4,5,6 expressions were not seen in human embryonic stem cells-derived neurogenesis, nor (B) *in vivo* in developing or adult brain. The gender of the embryo (M: male; F: female) and its age in weeks are indicated in the parentheses. Results are displayed as mean \pm standard error (s.e.m.).

Table S1. Oligonucleotide probes and conditions used for real-time PCR (qPCR) and semi-quantitative end-point RT-PCR

Gene	Primer sequences	Amplicon size	Tm	# cycles
hRFPL1 (qPCR)	F: CTg AAg gTg gTT CCC ATg TCT AT R: gAC TTg gAg gAg CAA AAA ACA AgT	81	-	-
hRFPL2-3 (qPCR)	F: CAg TgA TgA TTC gTg ACT TTC CCA R: TCC gAg AAC AAg ACC ATg AgT ACA	100	-	-
hRFPL4-6 (qPCR)	F: CAT TCA TCg AgA TTC CTg TTT gC R: CTg ATC ATC TTg ACT TCC ACg TTT	74	-	-
hRFPL1 promoter	F: ggg gAT gTg CTT gAg TgT TT R: ggA AAA gTg CAC AAg ggA Ag	148	57	40
hRFPL1 exon 2	F: gAA CAA AAT CAg gCC CAg TT R: CTT ggA ggA gCA AAA AAC AAg Tg	600	57	40
hRFPL2 promoter	F: CTA CTg CAT CCA gCC CCT CT R: CTg ACC AAC ATg ggg AAA C	238	57	40
hRFPL3 promoter	F: TCC TgT gTT gTC CAg ACT Cg R: gTg ggT ggA TCA TCT gAg gT	350	57	40
hRFPL2/3 exon 2	F: gAA CAA AAT CAg gCg CAA TC R: AAg gAg CCA AAA ATg ggC	596	57	40
murine Oct4	F: ggA Tgg CAT ACT gTg gAC CT R: AgA Tgg Tgg TCT ggC TgA AC	333	58	30
murine Nestin	F: CAg AgA ggC gCT ggA ACA gAg ATT R: AgA CAT Agg Tgg gAT ggg AgT gCT	469	61	32
murine β 3 tubulin	F: TCA gCg ATg AgC ACg gCA TA R: CAC TCT TTC CgC ACg ACA TC	301	60	30
murine Pax6	F: ggg ggT CTg TAC CAA CgA TA R: gTC CTC TCA AAC TCT TTC TC	369	58	30
mRFPL	F: gCC CTg gTT ACC AAg ATC AA R: gAT ATg ggA gCC ATC gCT AA	580	58	32
L32	F: gTg AAg CCC AAg ATC gTC AA R: TTg gTg ACT CTg ATg gCC Ag	349	58	26

Table S2. Identification of the RFPL gene family

Species	Gene name	chr	strand	Method of detection	start codon	stop codon	length	features
Human	hRFPL1	22	+	Refseq (NM_021026)	28,164,572	28,168,444	3873	
	hRFPL2	22	-	Refseq (NM_006605)	30,916,425	30,919,466	3042	
	hRFPL3	22	+	Refseq (NM_006604)	31,083,854	31,087,148	3295	
	hRFPL4	19	+	UCSC blat	60,980,742	60,982,099	1358	
	hRFPL5	19	+	UCSC blat	60,964,979	60,966,350	1372	
	hRFPL6	19	+	UCSC blat	60,974,983	60,976,354	1372	
Chimpanzee	cRFPL1	22	-/+	UCSC blat+Human-Chimp genomic alignment	28,278,755	28,285,339	-	Inversion of the first exon
	cRFPL2	22	-	ENCODE genomic alignment	31,087,967	31,091,008	3042	
	cRFPL3	22	+	ENCODE genomic alignment	31,241,462	31,244,742	3281	
	cRFPL4	19	+	UCSC blat+Human-Chimp genomic alignment	61,600,311	61,601,676	1366	
	cRFPL5	19	+	UCSC blat+Human-Chimp genomic alignment	61,546,683	61,548,053	1371	
	cRFPL6	19	+	UCSC blat+Human-Chimp genomic alignment	61,555,560	61,556,931	1372	
Orangutan	oRFPL1	22	+	UCSC blat	24,355,245	24,359,130	3886	
	oRFPL2	22	-	UCSC blat	27,200,319	27,202,725	2407	
	oRFPL3	22	+	UCSC blat	27,442,559	27,445,224	2666	
	oRFPL4	19	+	UCSC blat	57,598,445	57,599,839	1395	
	oRFPL5	19	+	UCSC blat	57,705,733	57,707,257	1525	Frame shift (173 AA only)
	oRFPL6	19	+	UCSC blat	57,620,267	57,621,670	1404	
Macaque	maRFPL1	10	+	UCSC blat+Human-Chimp genomic alignment	73,271,517	73,272,096	580	Only intron and exon 2; No ORF
	maRFPL2	10	-	ENCODE genomic alignment	76,103,734	76,106,122	2389	Frame shift (184 AA only)
	maRFPL3	10	+	ENCODE genomic alignment	76,266,920	76,267,499	580	Only intron and exon 2; No ORF
	maRFPL4	19	+	UCSC blat	61,808,701	61,810,062	1362	
	maRFPL5	19	+	ENCODE genomic alignment	61,784,808	61,786,175	1369	Intron similar to human RFPL4,5
	maRFPL6	19	-	ENCODE genomic alignment	61,792,106	61,793,454	1349	Intron similar to human RFPL4,5
Baboon	bRFPL2			ENCODE genomic alignment			2377	
	bRFPL3			ENCODE genomic alignment			580	Only intron and exon 2; No ORF
Marmoset	msRFPL1		-	UCSC blat	73,002	74,111	1110	contig 2246; only exon 1 and half intron; No ORF
	msRFPL4		+	UCSC blat	10,848	11,703	856	contig 22460; No ORF
	msRFPL5		-	UCSC blat	21,714	22,585	872	contig 5452; No ORF
	msRFPL6		+	UCSC blat	379,170	380,081	912	contig 996; No ORF
Mouse	mRFPL	7	-	Refseq (NM_138954)	5,061,468	5,068,498	7031	
Rat	rat RFPL	1	+	UCSC blat	67,466,536	67,471,161	4626	
Horse	horse RFPL	10	+	UCSC blat	19,719,133	19,721,529	2397	
Dog	dog RFPL	1	+	GeneWise Gene prediction	105,130,438	105,132,644	2207	
Cat	cat RFPL		+	UCSC blat	1648	3467	1820	scaffold_205584

BLAT Genome Assemblies used: Nov. 2004 (Rat), May 2005 (Dog), Jan. 2006 (Macaque), Mar. 2006 (Human, Chimpanzee, Cat), Jan. 2007 (Horse), June 2007 (Marmoset), July 2007 (Orangutan, Mouse).

Table S3. Calculations of K_S divergence between hominoid RFPL1,2,3 genes and between human and marmoset genomes (as derived in Nikolaev et al.²⁹)

Genes compared		K_S value
hRFPL1	hRFPL2	0.0501
hRFPL1	hRFPL3	0.0487
hRFPL2	hRFPL3	0.0256
cRFPL1	cRFPL2	0.0533
cRFPL1	cRFPL3	0.0521
cRFPL2	cRFPL3	0.0178
oRFPL1	oRFPL2	0.0157
oRFPL1	oRFPL3	0.0263
oRFPL2	oRFPL3	0.025
Human genome	Marmoset genome	0.109

Table S4. Parameter estimates and likelihood scores under models of variable K_A/K_S ratios (Ω) among sites of RFPL1-3 gene family

Nested model pairs	Average Ω of PSS	Number of PSS	PSS (%)	Likelihood	LRT	P value
background (branch-site)				-9433.184		
foreground (branch-site)	4.90687	28(13)	9.8	-9450.091	33.81343	p<0.001
M1 neutral				-9466.166		df=2
M2 selection	2.93946	19(8)	6.6	-9440.341	51.65018	p<0.001
M7 beta				-9463.333		df=2
M8 beta & omega	2.48823	26(13)	9.1	-9429.270	68.12613	p<0.001

LRT: Likelihood Ratio Test; PSS: positively selected sites with posterior probabilities >50% (>95%); df: degrees of freedom.

Table S5. Amino acid changes that occurred at sites inferred to have been potentially under positive selection in RFPL1,2,3 proteins (>95% with at least one model)

Residue	Domain	Model 2	Model 8	Branch-site
37	RING	0.956	0.989	
61	RING			0.952
83	RING			0.981
87	RING			0.997
89	RING			0.992
90	RING			0.996
93	RDM	0.937	0.985	
95	RDM			0.960
100	RDM			0.962
112	RDM	0.991	0.998	
113	RDM	0.803	0.951	
151	PRY	0.887	0.975	0.959
165	PRY	0.996	0.999	0.957
169	PRY			0.998
206	SPRY	0.973	0.993	0.776
208	SPRY	0.921	0.977	
213	SPRY	0.975	0.993	0.994
222	SPRY	0.994	0.998	
263	SPRY			0.953
282	SPRY			0.993
287	SPRY	0.960	0.990	0.805
289	SPRY	0.956	0.989	
304	SPRY	0.942	0.983	0.919

Residue positions are indicated according to hRFPL1 protein sequence. Note that some RFPL1,2,3 proteins have a N-terminal 29 amino acid deletion.