Table 1. Number of RP Pseudogenes on Each Chromosome

Chromo-some	Chr. size (Mb)	Chr. GC content	Functional RP genes	Processed Pseudogenes			Pseudogenic fragments	Processed + fragments	Pseudogene density [c]
				Intact [a]	Disrupted [b]	Total			
1	257	0.41	5	202	21	223	35	258	1.00
2	242	0.4	4	144	8	152	30	182	0.75
3	205	0.39	7	119	6	125	12	137	0.67
4	192	0.37	3	86	9	95	19	114	0.59
5	186	0.39	3	107	11	118	19	137	0.74
6	179	0.39	4	139	6	145	26	171	0.96
7	163	0.4	0	94	8	102	22	124	0.76
8	146	0.39	4	94	9	103	21	124	0.85
9	132	0.41	4	70	6	76	11	87	0.66
10	142	0.41	1	92	10	102	10	112	0.79
11	142	0.41	6	85	8	93	12	105	0.74
12	141	0.4	4	119	7	126	15	141	1.00
13	116	0.38	1	43	3	46	10	56	0.48
14	106	0.41	1	78	10	88	17	105	0.99
15	100	0.42	3	55	7	62	10	72	0.72
16	93	0.44	3	44	9	53	16	69	0.74
17	84	0.45	6	80	11	91	13	104	1.24
18	82	0.39	1	43	2	45	5	50	0.61
19	77	0.47	13	66	8	74	16	90	1.17
20	63	0.44	1	44	3	47	12	59	0.94
21	45	0.41	0	21	0	21	6	27	0.60
22	48	0.48	1	25	6	31	6	37	0.77
X	152	0.39	4	61	8	69	15	84	0.55
Y	59	0.39	1 [d]	1	2	3	0	3	0.05
Total	3152	0.41	80	1912	178	2090	358	2448	0.78

ᵃ Processed pseudogenes that are continuous in sequence with insertions $\leq$ 60 bp

ᵇ Processed pseudogenes that are disrupted by insertions (> 60 bp)

ᶜ Number of processed pseudogenes + pseudogenic fragments per 1 Mb DNA

ᵈ RSP4Y on chromosome Y is an isoform of RSP4X on chromosome X

Correlation (chromosome size, number of processed pseudogenes) = 0.89, P < 1E-8

Correlation (chromosome size, number of processed pseudogenes + fragments) = 0.89, P < 1E-8

Table 2. Overall Statistics of RP Processed Pseudogenes

	Ave. Relative sequence length [a]	Ave. a.a. identity [b]	Ave. nt. identity [c]	Ave. disablements [d]	Ave. 5' truncation [e]	Ave. 3' truncation [f]
Processed pseudogenes	96.5%	76.2%	86.8%	3.6	13	8
Pseudogenic fragments	42.1%	74.2%	84.1%	2.0	227	127
Total	88.5%	75.9%	84.8%	3.4	44	26

[a] Length of translated pseudogenes divided by the length of RP peptides, averaged over the entire pseudogene population.

[b] Average sequence identity between translated pseudogene and RP peptide sequence.

[c] Average sequence identity between pseudogene sequence and RP cDNA sequence.

[d] Average number of premature stop codons, frame shifts, and repeat insertions in the processed pseudogenes

[e] Average number of missing nucleotides at 5' end, CDS only.

[f] Average number of missing nucleotides at 3' end, CDS only.

Table 3. Genomic Distribution of RP Processed Pseudogenes

	Genomic GC-content [a]					
	< 37%	37-41%	41-46%	46-52%	>52%	Total
Number of functional RP genes [b]	8	6	23	27	22	86
RP gene density (per 100 Mb)	0.98	0.61	3.2	8.3	23.4	2.73
Number of RP pseudogenes	310	601	804	318	57	2090
Pseudogene density (per 10 Mb)	3.8	6.1	11	9.7	6.1	6.63

[a] Genomic regions grouped by their average GC content

[b] Including duplicated copies of the functional RP genes

Table 4. Distributions of Processed Pseudogenes Among RP Genes

Gene name [a]	SWISSPROT ID [b]	mRNA length [c]	CDS length [d]	CDS GC content [e]	# Processed	# Fragments
RPL21	RL21_HUMAN	568	483	0.43	145	13
RPL23A	RL2B_HUMAN	546	471	0.49	85	11
RPL7	RL7_HUMAN	838	747	0.43	83	19
RPL7A	RL7A_HUMAN	890	801	0.54	73	13
RPL31	RL31_HUMAN	442	378	0.47	71	5
RPSA	RSP4_HUMAN	1039	888	0.53	67	12
RPS26	RS26_HUMAN	459	348	0.53	65	5
RPS3A	RS3A_HUMAN	921	795	0.43	60	15
RPL17	RL17_HUMAN	898	555	0.46	59	11
RPS2	RS2_HUMAN	978	882	0.58	57	10
RPL39	RL39_HUMAN	401	156	0.42	56	8
RPL36A	RL44_HUMAN	425	321	0.47	54	6
RPL12A	RL12_HUMAN	632	498	0.52	49	3
RPL34	RL34_HUMAN	849	354	0.45	44	3
RPS15A	RS1A_HUMAN	541	393	0.47	43	4
RPL29	RL29_HUMAN	737	480	0.56	40	3
RPL26	RL26_HUMAN	525	438	0.45	39	3
RPS27	RS27_HUMAN	344	255	0.49	38	0
RPL5	RL5_HUMAN	1033	894	0.44	37	8
RPS20	RS20_HUMAN	539	360	0.45	36	0
RPL35A	R35A_HUMAN	511	333	0.47	36	2
RPL32	RL32_HUMAN	521	408	0.51	36	7
RPS29	RS29_HUMAN	346	171	0.53	36	2
RPS12	RS12_HUMAN	534	399	0.46	33	8
RPS10	RS10_HUMAN	598	498	0.55	33	4
RPL9	RL9_HUMAN	716	579	0.44	31	6
RPS24	RS24_HUMAN	537	402	0.44	30	13
RPL6	RL6_HUMAN	950	867	0.47	30	5
RPS6	RS6_HUMAN	829	750	0.47	27	1
RPL37	RL37_HUMAN	371	294	0.51	27	3
RPL13A	R13A_HUMAN	1142	612	0.58	26	3
RPS27A	R27A_HUMAN	551	471	0.44	24	1
RPL36	RL36_HUMAN	428	318	0.6	24	2
RPS4	RS4_HUMAN	916	792	0.48	23	5
RPL22	RL22_HUMAN	574	387	0.44	22	1
RPL19	RL19_HUMAN	698	591	0.54	22	4
RPL15	RL15_HUMAN	2018	615	0.54	21	5
RPS7	RS7_HUMAN	729	585	0.48	18	5

RPS17	RS17_HUMAN	515	408	0.51	18	3
RPP1	RLA1_HUMAN	512	345	0.54	17	3
RPL18A	RL1X_HUMAN	618	531	0.6	17	7
RPL30	RL30_HUMAN	524	348	0.45	16	1
RPS16	RS16_HUMAN	570	441	0.57	16	1
RPL10A	R10A_HUMAN	700	654	0.51	14	4
RPL10	RL10_HUMAN	2188	645	0.55	14	32
RPL18	RL18_HUMAN	648	567	0.59	14	1
RPS18	RS18_HUMAN	549	459	0.56	13	3
RPL13	RL13_HUMAN	1110	636	0.62	13	1
RPL23	RL23_HUMAN	493	423	0.49	11	1
RPL27	RL27_HUMAN	513	411	0.49	11	4
RPS8	RS8_HUMAN	705	627	0.52	11	3
RPP0	RLA0_HUMAN	1116	954	0.54	11	4
RPS28	RS28_HUMAN	398	210	0.62	11	2
RPS5	RS5_HUMAN	725	615	0.58	10	1
RPS15	RS15_HUMAN	515	438	0.63	10	1
RPS23	RS23_HUMAN	506	432	0.45	9	1
RPS25	RS25_HUMAN	514	378	0.47	9	0
RPL37A	R37A_HUMAN	392	279	0.52	9	1
RPL40 [f]	RL40_HUMAN	501	387	0.54	9	1
RPS21	RS21_HUMAN	356	252	0.54	9	0
RPL3	RL3_HUMAN	1311	1212	0.55	9	5
RPL35	RL35_HUMAN	455	372	0.57	9	2
RPS13	RS13_HUMAN	529	456	0.46	8	4
RPL24	RL24_HUMAN	556	474	0.48	8	9
RPS14	RS14_HUMAN	589	456	0.54	8	2
RPP2	RLA2_HUMAN	482	348	0.56	8	0
RPL38	RL38_HUMAN	368	213	0.46	7	4
RPS3	RS3_HUMAN	843	732	0.54	7	3
RPL4	RL4_HUMAN	1449	1284	0.49	6	5
RPS11	RS11_HUMAN	594	477	0.51	6	3
RPL27A [f]	RL2A_HUMAN	514	447	0.54	6	5
RPS19	RS19_HUMAN	569	438	0.58	6	0
RPL28	RL28_HUMAN	500	414	0.6	6	1
RPL41	RL41_HUMAN	478	78	0.5	5	0
RPL11	RL11_HUMAN	609	537	0.51	4	3
RPS30 [f]	RS30_HUMAN	574	402	0.59	4	1
RPS9	RS9_HUMAN	691	585	0.6	4	1
RPL8	RL8_HUMAN	894	774	0.61	4	3
RPL14	RL14_HUMAN	843	651	0.5	3	3

Total	685	511	0.51	26	4.5
Correlation	-0.01 [g]	0.04 [h]	-0.41 [i]		
p-value	0.93	0.73	0.0002		

[a] The SWISSPROT ID (Bairoch & Apweiler 2000) for the human RP protein

[b] The standard mammalian RP gene nomenclature

[c] Number of nucleotides for RP mRNA

[d] Number of nucleotides for RP coding sequence (CDS)

[e] GC content of the RP coding sequence

[f] RPS27a, RPL40 and RPS30 are carboxyl extensions of ubiquitin or ubiquitin-like proteins.

[g] Correlation between number of processed pseudogenes and mRNA length

[h] Correlation between number of processed pseudogenes and CDS length

[i] Correlation between number of processed pseudogenes and CDS GC-content

Table 5. Duplicated Human RP genes

gene	Duplicated			Original [a]		
	Chr. location	Size (bp)	Ensembl ID	Chr. location	Size (bp)	Ensembl ID
RPS3A	4q31.23	4114	ENSG00000145425	4q31.23	4114	ENSG00000151940
RPL26	5q35.3	9664	ENSG00000037241	17p13.1	4743	ENSG00000161970
RPL8 (I) [b]	8q24.3	2287	ENSG00000130795	8q24.3	2287	ENSG00000147785
RPL8 (II)	8q24.3	2287	ENSG00000161009	8q24.3	2287	ENSG00000147785
RPL7A	9q34.3	1814	ENSG00000160312	9q34.3	1814	ENSG00000148303
RLP0 (I) [c]	12q24.31	3474	N/A[d]	12q24.23	3474	N/A
RLP0 (II) [e]	12q24.31	2182	ENSG00000123062	12q24.23	3474	N/A
RPS27	15q21.3	912	N/A	1q21.3	573	ENSG00000157616
RPS17 [f]	15q26.3	685	ENSG00000154241	15q25.2	3329	ENSG00000103720
RPL3	16p13.3	8650	ENSG00000140986	22q13.2	5402	ENSG00000100316
RPS15A	16p13.11	6100	ENSG00000157115	16p13.11	6100	ENSG00000134419
RPL17	18q21.1	2221	ENSG00000141618	18q21.1	2220	ENSG00000154807
RPL13A [g]	19q13.13	1016	N/A	19q13.33	1966	ENSG00000142541
RPS9	19q13.42	6385	ENSG00000131036	19q13.42	6385	ENSG00000074164
RPL36	19p13.3	1108	ENSG00000130255	19p13.2	1108	ENSG00000141995
RPS4Y	Yq11.222	24252	ENSG00000157828	Yp11.31	24727	ENSG00000129824

[a] RP genes identified from hybridation experiments

[b c] Two copies of duplications are found for the same RP gene

[d] Not present in Ensembl. (Jan. 2002 release)

[e] This duplicated RP gene starts from amino acid 17

[f] This duplicated RP gene only contain amino acids 50-138.

[g] This duplicated RP gene has a frame shift

Table 6. Genomic Distributions of RP Pseudogenes of Different Ages

Sequence divergence	Number of pseudogenes in the genomic region [a]				
	<37%	37-41%	41-45%	45-52%	>52%
$\leq 1\%$	4	12	12	8	2
$\leq 2\%$	13	26	31	11	2
> 2%	303	575	768	306	55

[a] Genomic regions of 100 Kb long are binned by their average GC-content.