RESEARCH

A Map of 75 Human Ribosomal Protein Genes

Naoya Kenmochi,[1,3,4] Tomoko Kawaguchi,[1] Steve Rozen,[1,2] Elizabeth Davis,[1] Nathan Goodman,[2] Thomas J. Hudson,[2] Tatsuo Tanaka,[3] and David C. Page[1,2,4]

[1]Howard Hughes Medical Institute and [2]Center for Genome Research, Whitehead Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142 USA; [3]Department of Biochemistry, School of Medicine, University of the Ryukyus, Nishihara, Okinawa 903-01, Japan

We mapped 75 genes that collectively encode >90% of the proteins found in human ribosomes. Because localization of ribosomal protein genes (rp genes) is complicated by the existence of processed pseudogenes, multiple strategies were devised to identify PCR-detectable sequence-tagged sites (STSs) at introns. In some cases we exploited specific, pre-existing information about the intron/exon structure of a given human rp gene or its homolog in another vertebrate. When such information was unavailable, selection of PCR primer pairs was guided by general insights gleaned from analysis of all mammalian rp genes whose intron/exon structures have been published. For many genes, PCR amplification of introns was facilitated by use of YAC pool DNAs rather than total human genomic DNA as templates. We then assigned the rp gene STSs to individual human chromosomes by typing human–rodent hybrid cell lines. The genes were placed more precisely on the physical map of the human genome by typing of radiation hybrids or screening YAC libraries. Fifty-one previously unmapped rp genes were localized, and 24 previously reported rp gene localizations were confirmed, refined, or corrected. Though functionally related and coordinately expressed, the 75 mapped genes are widely dispersed: Both sex chromosomes and at least 20 of the 22 autosomes carry one or more rp genes. Chromosome 19, known to have a high gene density, contains an unusually large number of rp genes (12). This map provides a foundation for the study of the possible roles of ribosomal protein deficiencies in chromosomal and Mendelian disorders.

[The sequence data described in this paper have been submitted to GenBank. They are listed in Table 1.]

Although the ribosome, as catalyst for protein synthesis, is known to be essential for organismal growth and development, the effects of ribosomal mutations and their role in human disease have been explored barely. The mammalian ribosome is a massive structure composed of 4 RNA species and ~80 different proteins (Wool 1979). Typical mammalian cells contain about 4×10^6 ribosomes, and ribosomal RNAs and proteins constitute ~80% of all cellular RNA and 5%–10% of cellular protein. One might predict that genetic defects in ribosomal components would invariably result in early embryonic death. However, there is strong evidence in *Drosophila* and suggestive evidence in humans that quantitative deficiencies of ribosomal proteins may yield viable but abnormal phenotypes. In *Drosophila,* the *Minute* phenotype (reduced body size, diminished fertility, and short, thin bristles) results from heterozygous deficiencies (deletions) at any 1 of 50 loci scattered about the genome (Schultz 1929; FlyBase 1997). Several *Minute* loci have been characterized molecularly, and all have been found to encode ribosomal proteins (Kongsuwan et al. 1985; Hart et al. 1993; Andersson et al. 1994; Cramton and Laski 1994; Saebøe-Larssen and Lambertsson 1996; Schmidt et al. 1996; A. Cheng, A. Zinn, J. Mach, R. Lehman, and D.C. Page, unpubl.). Thus, it appears that reductions in the amount of any of a number of ribosomal proteins have a similar, characteristic effect on the development of *Drosophila* embryos.

Perhaps ribosomal protein deficiencies have analogous consequences in humans, resulting in specific, recognizable clinical features (which might or might not resemble the *Minute* phenotype observed in *Drosophila*). We and our colleagues have reported findings consistent with a role for ribosomal protein S4 (RPS4) deficiency in the etiology of

[4]Corresponding authors.
E-MAIL page@wi.mit.edu; FAX (617) 258-5578.
E-MAIL kenmochi@med.u-ryuku.ac.jp; FAX 81-98-895-7049.

certain features of Turner syndrome, a complex human disorder classically associated with a 45,X karyotype (Fisher et al. 1990; Watanabe et al. 1993; Zinn et al. 1994). We are intrigued by the possibility that deficiencies of other human ribosomal protein genes (rp genes) might cause phenotypic abnormalities similar to those seen in Turner syndrome—just as deficiencies of any of a number of *Drosophila* rp genes cause the *Minute* phenotype.

The ribosome is the largest, most complex mammalian structure to be completely described at the level of nucleotide and amino acid sequence. The nucleotide sequences of the four ribosomal RNAs—28S, 18S, 5.8S, and 5S—have been determined in their entirety (Maidak et al. 1997), and a systematic effort to deduce the primary structure of all mammalian ribosomal proteins by cDNA sequencing has come to completion (Wool et al. 1996).

Moreover, the genes encoding the RNA constituents of the mammalian ribosomes have all been assigned to chromosomes. The 28S, 18S, and 5.8S rRNAs are generated by elaborate processing of a single 45S precursor derived from tandemly repeated gene arrays which, in humans, are located on the short arms of chromosomes 13, 14, 15, 21, and 22 (Henderson et al. 1973; Worton et al. 1988). The 5S rRNA derives from tandemly repeated gene clusters on human chromosome 1 (Sørensen et al. 1991; Lomholt et al. 1995).

Paradoxically, only a small fraction of the genes encoding the mammalian ribosomal proteins have been mapped previously. Though these 80 proteins function together, their amino acid sequences are dissimilar. Unlike the ribosomal RNAs, each mammalian ribosomal protein typically is encoded by a single gene. However, in the case of most if not all ribosomal proteins, the single, functional gene has generated a large number of silent, processed pseudogenes at sites dispersed throughout the genome (Dudov and Perry 1984; Wagner and Perry 1985; Kuzumaki et al. 1987). These pseudogenes impede the mapping of the functional rp genes, explaining at least in part, why only 24 of the ~80 rp genes had been chromosomally assigned. The 24 genes that had been assigned map to 14 different chromosomes, suggesting that rp genes, unlike rRNA genes, are not clustered at a few sites in the genome (Feo et al. 1992).

If we are to explore systematically the possibility that ribosomal protein deficiencies or mutations cause certain human disorders, we must first learn the chromosomal map position of each of the ~80 human rp genes. This task is hindered by the existence of processed pseudogenes elsewhere in the genome. We developed general strategies to physically map human rp genes, while avoiding pseudogenes, using sequence tags specific to the functional, intron-bearing genes.

RESULTS

The human rp genes had been previously characterized to varying degrees. Some human rp genes had been completely sequenced at both the cDNA and genomic levels, whereas for others, even partial cDNA sequences were unavailable. We divided the estimated 80 human rp genes into three classes, ordered according to how much gene structure information was available (Table 1). For group 1, containing 19 genes, at least some human splice sites had been determined and some human intron sequence was available. No human splice site information was available for any of the remaining 61 rp genes. However, for 12 of these 61 genes, at least some splice sites had been determined in the homologous genes in rat, mouse, chicken, or frog. As described below, the extreme conservation of splice-site positions among homologous vertebrate rp genes allowed us to predict the positions of splice sites in the human genes using this information. These 12 rp genes, for which somewhat less information was available, constitute group 2. No vertebrate splice-site information was available for any of the remaining 49 rp genes, which comprise group 3.

For each of the three groups, we developed a separate strategy for generating sequence tags specific to the functional, intron-bearing genes. For all three groups, we exploited the fact that rp pseudogenes, derived from processed transcripts, lack the introns found in their progenitors (Dudov and Perry 1984; Davies et al. 1989). For group 1 genes, identification of STSs was straightforward. We derived STSs specific to the functional genes by choosing oligonucleotides from the previously sequenced introns (Fig. 1A).

STSs for Group 2 and 3 Genes: Intron Trapping

For group 2 and group 3 genes, no human intron sequences were available. However, other information allowed us to predict the location of, and then trap, introns from these genes. The methods we employed were inspired in part by earlier successes in identifying and mapping intron-bearing rp genes by PCR (Davies et al. 1989).

For group 2 genes, the positions of at least some splice sites in vertebrate homologs were known, and this information played a central role in our map-

Table 1. Chromosomal Assignments and STSs for Human Ribosomal Protein Genes

Gene	Chromosome	Group	Forward primer[a]	Reverse primer[a]	PCR prod. size (bp)	STS name	Accession no.[b]
RPSA	3	3	GGAGGAATTTCAGGGTGAAT	AGACCAGTCAGTGGTTGCTC	426		J03799
			GGAGGAATTTCAGGGTGAAT	ccactcccaacagcagtcta	272	RP_SA_1	AB007146[d]
RPS2	16	3	CTGCTCATGATGGCTGGT	CACAGCTGGAGCCTGAGT	460		X17206
			AACTTCGgtaggtggtccac	caggagggtcagtggtgtg	214	RP_S2_1	AB007147[d]
RPS3	11[e]	1	catgtggtcccacctattcc	gggggaaaagtgacaattca	263	RP_S3	L16016
RPS3A	4[f]	3	CCTCTCCATGATGTCTTCGT	TTCTTGGACTGGTGGTTCAT	358		M77234
			aaatcacatgattcctgtaggg	GCTTTCCCActacaaggcaa	187	RP_S3A_1	AB007148[d]
RPS4X	X[g]	1	tgagatggattgaatgtggc	ttaaagagggtgcccaggta	249	RP_S4X	AF041428
RPS4Y	Y[g]	1	ttaaggggacagtatttcaacttc	ccacatttaaactgagtacagtcc	361	RP_S4Y	AF041427
RPS5	19	3	CATCTGGCTGCTGTGCAC	GTTGGACTTGGCCACACGCT	246		U14970
			CTCATCAATGCTGCCAAGgt	ccacaaatgcccaacttaag	66	RP_S5	AB007149[d]
RPS6	9[h]	1	aacgccattggtaactggta	agtcgcatttctaaccgatg	325	RP_S6	X67309
RPS7	2[i]	1	aggttcagccacagtgagag	acccaccccctaactaccaag	249	RP_S7	Z25749
RPS8	1[j]	1	tctcagttcctttgactgcc	atctctgtgggttcaggtacc	307	RP_S8	X67247
RPS9	19	3	TCGCAAAACTTATGTGACCC	CTTCGAACAGACGCCGTG	399		U14971
			agtccgtcccctaaatttgg	CTTCGAACAGACGCCGTG	245	RP_S9	AB007150[d]
RPS10	6	3	CCGGATTGCCATCTATGAAC	GCAGATGAAGGTAATCACGG	468		U14972
			atggagggatcttcccctaa	GCAGATGAAGGTAATCACGG	266	RP_S10_2	AB007151[d]
RPS11	19[k]	3	CTATCTGCACTACATCCGCA	CTTCTGGAACTGCTTCTTGG	1663		X06617
			ttccagatcggaccaattta	tactcgccaattcaacctgt	1287	RP_S11	AB007152[d]
RPS12	6	3	CTTTTGGCTGAACACCAAAT	TCTTCAATGACATCCTTGGC	1100[c]		X53505
			gaagtcgttaaggtcctgag	TCTTCAATGACATCCTTGGC	206	RP_S12	AB007153[d]
RPS13	11	3	ATGCTAAATTCCGTCTGATTC	GGCAGAGGCTGTAGATGATT	745	RP_S13	L01124
RPS14	5[l]	1	ttccattagcctgccttttg	ccccagaaactaccaacagc	388	RP_S14_1	M13934
RPS15	19	1	agacgagtttcagtgttgcc	agaccacagcctcagacaag	319	RP_S15	M32405
RPS15A	16	3	CATGAATGTCCTGGCAGATG	CATCAATGATTTCAAATTCGC	900[c]		X84407
			ttgttttgggtattgtctccc	CATCAATGATTTCAAATTCGC	236	RP_S15A_1	AB007154[d]
RPS16	19	2	GGTGGCCTATTACCAGAAATgt	GGAAGCCTCATCCACATct	130[c]	RP_S16_1	M11408,M60854
RPS17	15[z]	1	tcacccagctagtaagtggc	gtggttagcatggagacacc	281	RP_S17	M18000
RPS18	6	2	GATGGAAAATACAGCCAGgt	GCCTCGGACACGAAGGct	350[c]	RP_S18	M76762,X69150
RPS19	19	3	AGATCTATGGGGGACGTCAG	CTGTCCGGCGATTCTGT	650[c]		M81757
			acctgactagggccctcagt	CTGTCCGGCGATTCTGT	191	RP_S19_1	AB007155[d]
RPS20	8	3	TGAAAGGACCAGTTCGAATG	TCAACTCCTGGCTCAATACTG	607		L06498
			TGAAAGGACCAGTTCGAATG	caccaagaacacagcaacaa	233	RP_S20	AB007156[d]
RPS21	20	3	CAATCGCATCATCGGTGC	CTTGGCCAATCGGAGAAT	723		L04483
			ggtgaaggtacaggcagagg	gatgtcccgaagtgatgagc	199	RP_S21	AB007157[d]
RPS23	5	3	CTTTGTACCCAATGACGGTT	ATCTTGGTCTTTCCTTCTTGC	324		D14530
			ttctgttgttggggtggc	ttatttctggtgttggtggc	101	RP_S23_1	AB007158[d]
RPS24	10[m]	2	CCCAAACATAGACTTGCAAGAgt	CTTCTCATACAGGCCATGct	1300[c]		X71972,M31520
			gggaaagaccaagcaatctg	ccaggctacctttccaaaca	217	RP_S24_2	AB007159[d]
RPS25	11[n]	1	AGGAGCTCCTTAGTAAAGgtgag	TGAAACCAGCTTGATAAGTCct	1200[c]	RP_S25	D15048,M64716
RPS26	12	3	AAATTCGTCATTCGAAACAT	AGCACCCGCAGGTCTAAATC	900[c]		X69654
			AAGCGAGCGTCTTCGATGgt	GCTTGGGAAGCACATAGGct	750[c]	RP_S26	AB007160[d],AB007161[d]
RPS27	1	3	GATCTCCTTCATCCCTCTCC	GTAAGCCTTGCTTTTCCTCC	550		L19739
			cgtatccttgaagctgtgca	gtttcccaacacaaaaaaacg	184	RP_S27_2	AB007162[d]
RPS27A	2	3	GTTAAGCTGGCTGTCCTGAA	GTAAGTCAGACAACATTTGCCA	618		X63237
			GTTAAGCTGGCTGTCCTGAA	tttccagaatacctcatttaaacc	169	RP_S27A_1	AB007163[d]
RPS28	19	3	CGACGTGCTCACCCTTTTG	GAGCATCTCAGTTACGTGTGG	342		L05091
			ttgatagacctttggttggtg	acatgagatgttgacaggcc	141	RP_S28_1	AB007164[d]
RPS29	14	3	ACGGTCTGATCCGGAAATAT	AAGAATTATCATGGTTTTTCATTG	2300[c]		U14973
			ACGGTCTGATCCGGAAATAT	aaagggagtaggtaagactaagtcc	310	RP_S29	AB007165[d]
RPS30	11[o]	1	aagcgtctagtgagtgtggg	ggcctggtctgacctaattt	231	RP_S30	X65921
RPL3	22	3	CCAATGACTTTGTCATGCTG	TTGCAATTCGGTCTTTCTTC	745		X73460
			ggcaggtcctgacttgtcat	TTGCAATTCGGTCTTTCTTC	673	RP_L3_1	AB007166[d]
RPL4	15	2	GCTTCCCTCAAGAGTAACTACAAgt	ATCTTGTGCATGGGAAGAct	600[c]		M15678,D23660
			GCTTCCCTCAAGAGTAACTACAAgt	gtggtggctcacccctataa	400[c]	RP_L4_2	AB007167[d]
RPL5	1	2	CAAACAGAGATATCATTTGTCAGgt	TCTATACGGGCATAAGCAATct	1200[c]	RP_L5	D10737,U14966
RPL6	12	3	CCCAAAACATCTTACTGATGC	GATAAATTCCATTCGTCAGAGC	700	RP_L6	D17554
RPL7	8	1	ggtcagtgccttgcagtagt	ctcccgtactgagcagtgtt	296	RP_L7	L16557
RPL7A	9[p]	1	atgcttcttgctttcattgg	cactaggaagggtgttggtg	299	RP_L7A	X52138
RPL8	8	3	TGATCCGTGGACAGAGGAAG	TGGCATAGTTCCCTGATGC	747		Z28407
			gaggagttaacgtcctaaaccg	gcagtatcaggccggcag	206	RP_L8_1	AB007168[d]
RPL9	4[q]	3	CATTGAGCTTGTTTCAAATTCA	TCATCTGGCATCTTCTTTCTG	693		D14531
			aatggagagttaagcattctgg	cttgcccaggaaagtactaatc	170	RP_L9_1	AB007169[d]

(*Continued on following page.*)

Table 1. (*Continued*)

Gene	Chromosome	Group	Forward primer[a]	Reverse primer[a]	PCR prod. size (bp)	STS name	Accession no.[b]
RPL10	X[r]	3	TTATCATGTCCATCCGCACC	TGGACAGGAAGTAGAATTTATTGG	379		M73791
			TTATCATGTCCATCCGCACC	ggagcaatggaaagaactga	163	RP_L10	AB007170[d]
RPL10A	6	1	gcaccgttctgatcccac	CCCAGGACACACACAGAGAA	395	RP_L10A	U12404,Z62020
RPL11	1	3	ATCCGCAAACTCTGTCTCAA	GTGCAGTGGACAGCAATCTT	1000[c]		L05092
			ATCCGCAAACTCTGTCTCAA	gactcgacagcaggtaaatca	187	RP_L11_1	AB007171[d]
RPL12	9	3	ATTGTCAACATTGCTCGAC	GCATTCCACAGCACCACT	509	RP_L12	L06505
RPL13	16[s]	3	AACTGAAACTGGCCACCC	TTCAACATCCTGTTCTGCG	689		X64707
			CATGCCCGTCCGGAACgt	CGAGCTTTCTCCTTCTTATAGACcta	531	RP_L13	AB007172[d]
RPL13A	19	2	GCTAAACAGGTACTGCTGGgt	AGGAAAGCCAGGTACTTCAct	592	RP_L13A	X51528,X56932
RPL15	3	3	TCTGATGTCATGCGCTTTCT	GTTAACACCATGATGGACAGG	600[c]		L25899
			TCTGATGTCATGCGCTTTCT	agagagacactgaggcagca	223	RP_L15_1	AB007173[d]
RPL17	18	3	CATGGTCGGATTAACCCATA	CCGTGCCATAAGTTTTTGTT	900[c]		X53777
			ccccacttagatgtacatagcc	tggaggacttcagcttattctg	235	RP_L17	AB007174[d]
RPL18	19	2	CGCACAGCCACACCAAgt	GGAGCGGACGTAGGGTct	400[c]	RP_L18	X05025,L11566
RPL18A	19	3	CCCACTCCATTCAGATCATG	TTTATTTGGGGCACACCC	348		L05093
			ctcccctggagggaagtg	aggcatgttgaggggtt	113	RP_L18A	AB007175[d]
RPL19	17[t]	2	AGATCGATCGCCACATgt	TCAGGTACAGGCTGTGATACct	900[c]	RP_L19	M30264,X63527
RPL21	13	3	ACCTGGGTTCAACTAAAGCG	GAATAGGTTCCAGCAGCTCA	195		U14967
			ACCTGGGTTCAACTAAAGCG	GGTGGAGCAGGctattaaagg	136	RP_L21_1	AB007176[d]
RPL23	17	3	TGGTCATTCGACAACGAA	CAGCATTGGATGCAATCC	318	RP_L23	X52839
RPL23A	17[u]	1	atacattacccgcccctctc	ggtgcatcagtggttccttt	231	RP_L23A	AF001689
RPL24	3	3	TGGTGCATCTCTTGCTGATA	TTAGCGTTTTCCACCAACTC	1600[c]		M94314
			tctggatgtgagagatgggt	TTAGCGTTTTCCACCAACTC	155	RP_L24_1	AB007177[d]
RPL26	17	1	cgttgattgtaccttctaaaacg	tccaaacaaataaccgcaat	344	RP_L26	L07287
RPL27	17	3	CCTTGGACAAAACTGTCGTC	CCACTTGTTCTTGCCTGTCT	207	RP_L27	L19527
RPL27A	11	3	GTGAACCTTGACAAATTGTGG	CCCCCCAACACTCTTAATCT	1000[c]		U14968
			GTGAACCTTGACAAATTGTGG	taggcactgctacggatagg	267	RP_L27A_1	AB007178[d]
RPL28	19	3	ACCTCCTATGTGCGGACCAC	CTCTTCACCATCACAGGCTT	356		U14969
			ACCTCCTATGTGCGGACCAC	atcccctctccagcccatatt	198	RP_L28_1	AB007179[d]
RPL29	3[v]	1	GCCAAGTCCAAGAACCA	tgtgggagaaaaaaagggaa	362	RP_L29	U10248
RPL30	8[k]	2	GGCCGCAAAGAAGACGgt	CGACTCCAGCGACTTTTTct	280[c]	RP_L30	D14521,L05095
RPL31	2	3	GAATGTGCCATACCGAATCC	TTAGTTCTCATCCACATTGACTG	424		X15940
			agttctccatcccataaagcc	ccttttgtagacgggtctgg	212	RP_L31_1	AB007180[d]
RPL32	3	2	AGACCGATATGTCAAAATTAAGgt	GTTTCCGCCAGTTACGct	650[c]	RP_L32	K02060,X03342
RPL34	4	3	CGTTTGACATACCGACGTAG	CACACACCACATGCAGATTT	217		L38941
			caaaatgctgacctactgactg	CACACACCACATGCAGATTT	116	RP_L34_1	AB007181[d]
RPL35A	3[w]	3	CGGCAAACCAAACAAAAC	TCCTTGAGGGGTACAGCA	1800[c]		X52966
			aagtgatgggattacaggtgtg	TCCTTGAGGGGTACAGCA	246	RP_L35A	AB007182[d]
RPL36A	X[x]	1	agagcaacccaatcttgcc	TGCTTGCCACACTTCTTACA	353	RP_L36A_1	L35265
RPL37	5	3	CGACCTGTGGCAAATGTG	TTAAGATGAACTGGATGCTGC	1800[c]		L11567
			CCGCAGATTCAGgtacagtt	TTAAGATGAACTGGATGCTGC	1500[c]	RP_L37_1	AB007183[d],AB007184[d]
RPL37A	2	2	GGTGCCTGGACGTACAAgt	GTGACAGCGGAAGTGGTAct	1200[c]	RP_L37A	D14167,X66699
RPL38	17	3	AGTTTAAAGTTCGATGCAGCA	TCATTTCAGTTCCTTCACTGC	611	RP_L38_1	Z26876
			tcgaggttggtgctagtactg	TCATTTCAGTTCCTTCACTGC	281		AB007185[d]
RPL40	19[y]	1	gtgtcagtctcagactcccc	agaaggacggcgactgag	220	RP_L40	X56997
RPL41	12	3	ACTTCGCCTTTCTCTCGG	CTTCTTTTGCGCTTCAGC	879		Z12962
			ccgccatacctcctgaacta	ggaacttcccctcctcctgg	380	RP_L41_1	AB007186[d]
RPP0	12	3	ATCATCAACGGGTACAAACG	ATATCCTCGTCCGACTCCTC	601		M17885
			atccacaccaggaccacagt	agcctttcctgtcagaagca	271	RP_P0_1	AB007187[d]
RPP1	15	3	TGGCCAACGTCAACATTG	GTCATCATCAGACTCCTCGG	310	RP_P1	M17886
RPP2	11	2	GACGTCATTGCCCAGGgt	GCAAGCTTGCCAATACct	850[c]	RP_P2	X55153,M17887

[a]Upper- and lowercase letters denote exon and intron sequences, respectively.
[b]Sequence from which STS was generated.
[c]Size (in bp) estimated by agarose gel electrophoresis.
[d]Sequence determined in this study.
[e–y]Previously reported chromosomal assignments were confirmed: [e]Polakiewicz et al. 1995; [f]Nolte et al. 1996; [g]Fisher et al. 1990; [h]Antoine and Fried 1992; [i]Annilo et al. 1995; [j]Davies and Fried 1993; [k]Feo et al. 1992; [l]Rhoads et al. 1986; [m]Jones et al. 1997; [n]Imai et al. 1994; [o]Kas et al. 1993; [p]Yon et al. 1993; [q]Mazuruk et al. 1996; [r]Van den Ouweland et al. 1992; [s]Adams et al. 1992; [t]Davies and Fried 1995; [u]Fan et al. 1997; [v]Garcia-Barcelo et al. 1997; [w]Colombo et al. 1996; [x]Oeltjen et al. 1995; [y]Webb et al. 1994.
[z]Previously reported chromosomal assignment (Feo et al. 1992) was corrected.

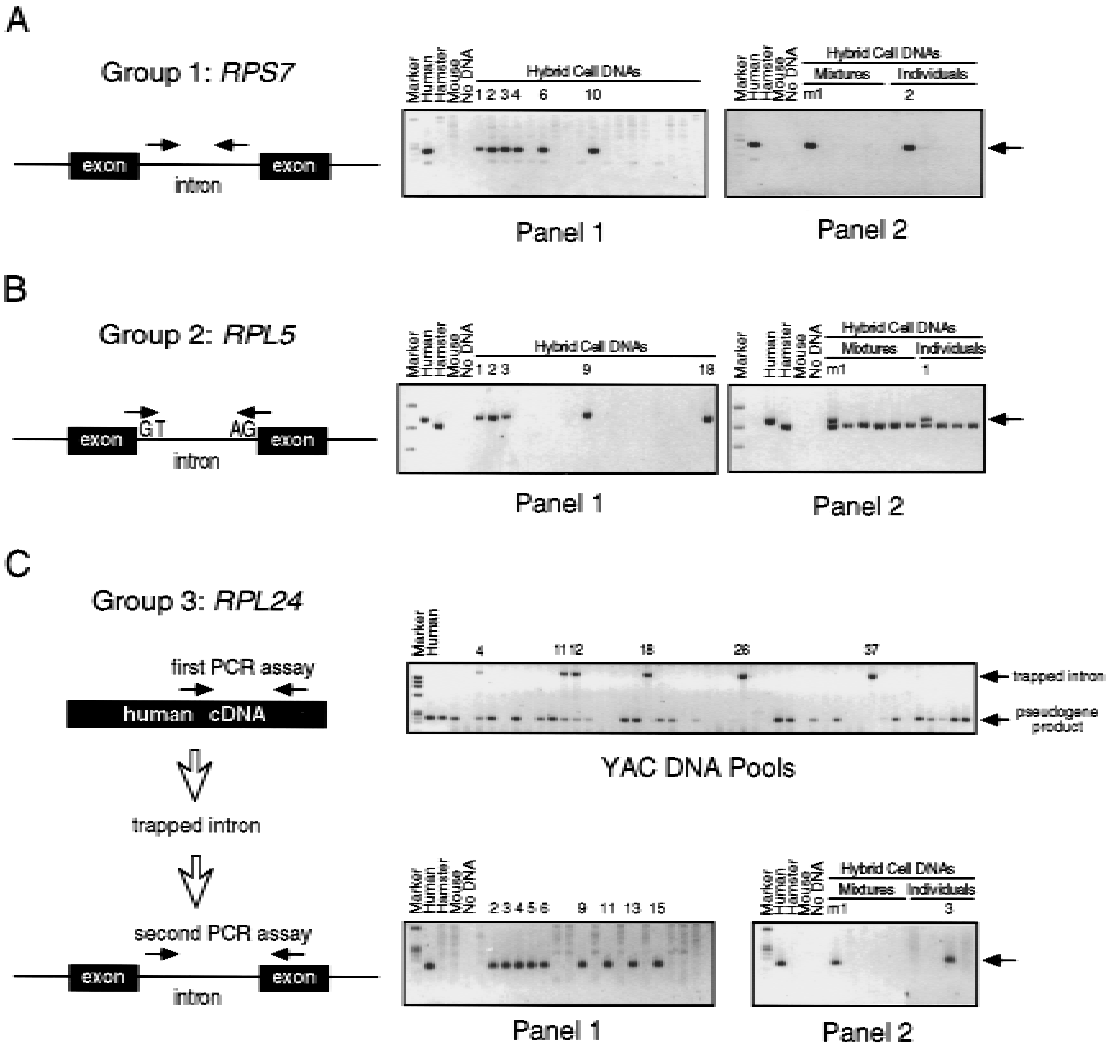

Figure 1 Intron trapping: derivation of STSs for and chromosomal assignment of representative group 1, group 2, and group 3 rp genes. (*A*) (Group 1) PCR primers were chosen from a previously sequenced intron of human *RPS7*. (*Right*) Results of testing of human/rodent somatic cell hybrid DNAs (NIGMS panels *1* and *2*) for the *RPS7* STS; agarose gel stained with ethidium bromide. On each gel, the left-most lane contains size markers, and the next four lanes show results of PCR controls with human, hamster, or mouse genomic DNA, or no added DNA, as template. (Panel *1*) 18 hybrid lines, each retaining multiple human chromosomes; 6 hybrids that tested positive for *RPS7* are numbered; the results map *RPS7* to chromosome 2. (Panel *2*) 24 hybrid lines, most retaining a single human chromosome; we initially tested these hybrids in pools of four (here mixture m1 is positive) and then tested individual hybrids from the positive pool (the chromosome 2 hybrid is positive). (*B*) (Group 2) PCR primers were chosen from human *RPL5* cDNA sequence at a predicted splice site, with the dinucleotide GT appended to the 3′ end of the forward primer and the dinucleotide CT (the complement of AG) appended to the 3′ end of the reverse primer. (*Right*) Results of hybrid mapping, which assigned *RPL5* to chromosome 1. The arrowhead (extreme *right*) indicates the size of the human PCR product; a smaller, hamster product is also present in many lanes. (*C*) (Group 3) Mapping of *RPL24* involved two quite different PCR assays. The first PCR assay (*top*), with 45 YAC DNA pools as template, was designed to trap an intron with primers chosen from human cDNA sequence according to rules discussed in the text. Six YAC pools that yielded the higher molecular weight, trapped-intron product are numbered; many more pools yielded the lower molecular weight, pseudogene product. The control reaction with human genomic DNA as template yelded only the pseudogene product. Sequencing of the trapped intron made possible a second PCR assay specific to the functional RPL24 gene (*bottom*); with the second assay, the gene was mapped to chromosome 3.

ping strategy. A survey of all rp genes whose intron/exon structures had been determined in any two vertebrates revealed universal conservation of intron location but little conservation of intron sequence (analysis not shown; for examples, see Maeda et al. 1993; Annilo et al. 1995; Davies and Fried 1995). Thus, we could predict the sites of introns within human group 2 cDNA sequences by comparison with more thoroughly studied vertebrate homologs. These predictions enabled us to avoid coamplification of pseudogenes by choosing PCR primers that were likely to contain splice sites and to extend slightly into introns. We exploited the fact that introns usually have a GT dinucleotide at their 5′ end and an AG dinucleotide at their 3′ end. As diagrammed in Figure 1B, the forward and reverse primers were based on human cDNA sequence immediately preceding and following a predicted splice site, but with the dinucleotide GT appended to the 3′ end of the forward primer and the dinucleotide CT appended to the 3′ end of the reverse primer. Thus, both the forward and reverse primers extended two nucleotides into the intron. The resulting PCR product was a trapped intron.

We avoided splice sites where the human cDNA sequence is AG/GT, as in such cases processed pseudogenes would be expected to be perfect templates. In a few cases where primers immediately flanking one splice site had markedly different melting temperatures or were otherwise predicted to be incompatible, primers were selected from two consecutive splice sites (i.e., primers predicted to span the outer borders of two consecutive introns).

For group 3 genes, we had no prior information about splice sites in any vertebrate. These genes could not be mapped by use of the group 1 or group 2 methods. However, if we could trap rp gene introns using human cDNA sequence as our only starting information, we might then derive intron-based sequence tags that would identify the functional genes as distinct from their pseudogenes. We arrived at the scheme diagrammed in Figure 1C. Using a forward primer likely to derive from the penultimate exon and a reverse primer likely to derive from the final exon, we attempted to PCR amplify the final intron of each group 3 gene. The details of the strategy emerged from several considerations. We would have to confront not only cross-amplification of pseudogenes but also the possible failure of the functional gene to amplify because the sequence complementary to one (or both) of the PCR primers was interrupted by an intron. We noted that, for most human rp genes whose intron/exon structure has been determined, the 50 nucleotides upstream of the termination codon are not interrupted by an intron. Thus, we chose the reverse primer from within a 50-bp region immediately upstream of the termination codon; this reverse primer was likely to derive from the final exon. We also noted that the terminal exon of human rp genes rarely extends >150 bp upstream of the termination codon, and that internal exons average 100 bp in length. Thus, we chose the forward primer from within the region 150–250 bp upstream of the termination codon; this forward primer was likely to derive from the penultimate exon or the one preceding. Further, we noted that about half of all splice sites in human ribosomal protein coding sequences are preceded by the dinucleotide AG. To reduce the probability that primers would span splice sites, we chose forward primers that did not include the sequence AG.

The rules just outlined were intended to (1) maximize the probability that PCR primer pairs would flank one or two introns (products containing three or more introns might be more difficult to amplify) and (2) minimize the probability that either primer would span a splice site. However, these rules for primer selection would not prevent amplification of closely related pseudogenes. We were concerned that during PCR amplification of human or human–rodent hybrid genomic DNA, competition between functional gene and pseudogene templates would occur; such competition would usually favor the pseudogenes because of their greater number, their high sequence similarity to the functional genes, and the smaller size of the resulting PCR product. In an effort to circumvent this problem, we tested PCR primer pairs on 45 pools of random human YACs, each pool comprising a random fifth of the human genome (Chumakov et al. 1995). (The YAC inserts averaged 0.6–1.0 Mb, and each pool contained 768 YACs.) We reasoned that among the 45 YAC pools, 9 or 10 should contain the functional gene and that, of these, one or more pools might be relatively free of competing pseudogenes. Such YAC pool roulette often yielded a jackpot (a clearly discernible, higher molecular weight PCR product, presumably containing one or more introns) in one or more YAC pools, even when use of the same primers on human genomic DNA yielded only pseudogene products (e.g., *RPL24,* as shown in Fig. 1C).

We then sequenced the higher molecular weight products obtained by PCR on YAC pools to confirm the presence of a genuine rp gene intron flanked by consensus splice sites. Indeed, by sequencing such PCR products, we identified a total of 55 new splice sites within 44 of the group 3 genes

(Table 2). For 34 of these genes, the PCR product contained a single intron, in which case we identified a single splice site. For the other 10 genes, the PCR product contained two (or, in one case, three) introns, in which case we identified two (or three) splice sites. All splice junction sequences conformed to the GT/AG rule and approximated the larger consensus sequence described by Mount (1982). The trapped introns varied in length from 75 to ~2100 bp. With the sequence of the trapped introns in hand, we were then able to design new PCR assays that amplified the functional rp genes but not their processed pseudogenes. These new, intron-based STSs were employed in subsequent mapping experiments.

Chromosomal Assignments

In all, we succeeded in identifying PCR-assayable STSs for 75 rp genes, including all 19 group 1 genes, all 12 group 2 genes, and 44 of 49 group 3 genes. To assign each of these 75 rp genes to an individual human chromosome, we tested two panels of human–rodent hybrid cell line DNAs (Drwinga et al. 1993) for the presence of the corresponding STS. For each of the rp gene STSs the chromosomal assignments derived by use of the first and second hybrid panels were concordant. (Three examples are shown in Fig. 1.) In this manner, each of the 75 rp genes was unambiguously mapped to a single human chromosome (Table 1). Chromosomal assignments had been reported previously for 24 of these 75 genes, and in 23 cases we confirmed these prior studies. In only one case (*RPS17,* on chromosome 15) do our results contradict a previous assignment.

Fine Localization

We employed two methods, radiation hybrid (RH) mapping and YAC/STS content analysis, to localize more precisely the 75 chromosomally assigned rp genes. For RH mapping, we scored for the presence or absence of each of the rp gene STSs in 91 human–hamster hybrid cell lines comprising the GeneBridge 4 whole-genome RH panel (Walter et al. 1994). This RH panel had been used previously to construct a comprehensive, STS-based map of the human genome (Hudson et al. 1995). Analysis of the rp gene STS typing results allowed us to position 73 of the rp genes on this pre-existing map (Fig. 2).

In parallel, we attempted to place the chromosomally assigned rp genes on a pre-existing YAC/STS content map of the human genome (Hudson et al. 1995). In this case, the CEPH YAC library (Chumakov et al. 1995) was screened by PCR to identify clones containing human rp genes. We identified a total of 222 YACs that carry 55 different rp genes. Information as to the identities, STS content, and chromosomal location of these 222 ribosomal protein-encoding YACs is available at an Internet site maintained by the Whitehead Institute/MIT Center for Genome Research (http://www-genome.wi.mit.edu/cgi-bin/contig/phys_map).

By integrating the results of our RH and YAC mapping efforts, we were able to place all 75 chromosomally assigned rp genes on the map of the human genome (Fig. 2). Regional assignments had been reported previously for some of these genes; in only one case (*RPL29,* on chromosome 3p) do our results conflict with a previous regional localization.

DISCUSSION

The ribosome is the most complex mammalian structure to have been completely described at the level of nucleotide and amino acid sequence (Wool 1996; Maidak et al. 1997). We now know the chromosomal locations, in humans, of the genes encoding all 4 RNAs and 75 of an estimated 80 proteins comprising this elaborate, protein synthesis machine (Fig. 2). Most mammalian ribosomal proteins have recognized homologs in prokaryotes, where rp genes are organized into a small number of operons, with as many as 11 ribosomal proteins under the control of a single promoter (Nomura et al. 1984). In contrast, there is little evidence of rp gene clustering in mammals (Feo et al. 1992)—a conclusion that our results confirm and extend. Both human sex chromosomes and at least 20 autosomes (all but chromosomes 7 and 21) carry one or more rp genes (Table 1; Fig. 2). Only the presence of 12 rp genes on chromosome 19, which constitutes only two percent of the haploid genome (Morton 1991), is notably at odds with a random distribution of rp genes throughout the human genome. Chromosome 19 is known to have a high gene density (Schuler et al. 1996), and even here, the 12 rp genes are scattered. With 55 rp genes mapped to YAC clones, we found only two examples of multiple rp genes residing on the same YAC clone: *RPS26* and *RPL41* (on chromosome 12), and *RPS11* and *RPL13A* (on chromosome 19). If one considers both ribosomal RNAs and proteins, it is apparent that virtually every human chromosome (except perhaps chromosome 7) contributes one or more components to the ribosome (Fig. 2). Though encoded at dispersed genomic sites, the ribosome's myriad components are apparently as-

sembled with stoichiometric precision. Regulated coproduction of the components could, in theory, be achieved in several ways. The clustering of rp genes in operons, as in bacteria (Nomura et al. 1984), is evidently not an important means of regulated coproduction in humans. *Trans*-acting regulatory mechanisms, both transcriptinal and translational, have been argued to play a substantial role in coordinating production of ribosomal components in mammals (Hariharan et al. 1989; Meyuhas et al. 1996), though feedback mechanisms, if any, remain to be elucidated. Alternatively, some ribosomal components may simply be produced in excess, with molecules not incorporated into ribosomes being discarded.

A few human rp genes remain to be mapped. In five cases (*RPL14, RPL22, RPL35, RPL36,* and *RPL39*), we were unable to trap verifiable introns despite repeated efforts; we did not map these five genes. We anticipate that these genes will be mapped as more information about their gene structures becomes available. [One of these genes, *RPL22,* has been assigned to chromosome 3 (Nucifora et al. 1993).] As a rule, each mammalian ribosomal protein is encoded by a single functional gene (Dudov and Perry 1984; Wagner and Perry 1985; Kuzumaki et al. 1987), but we cannot exclude the possibility that a second functional gene may exist in some cases. Indeed, functionally interchangeable isoforms of RPS4 are encoded by the human X and Y chromosomes (Fisher et al. 1990; Watanabe et al. 1993). In the case of *RPL36A,* a functional, intron-bearing gene is located on the X chromosome (Oeltjen et al. 1995), but analysis of cDNA sequences suggests that a second functional gene may exist elsewhere in the genome, as yet unmapped (N. Kenmochi et al., unpubl.).

Table 2. Newly Identified Splice Sites in Human Ribosomal Protein Genes

Gene	Exon		Intron size (bp)		Exon
	830[a]				
RPSA	TTCCCTACTG	gtatgtatca	- - 212 - -		ctcttaacag	AAGACTGGAG
	733				
RPS2	GGCAACTTCG	gtaggtggtc	- - 226 - -		tgttttgcag	CCAAGGCCAC
	690				
RPS3A	AAGTTTGAAT	gtaagtgaga	- - 187 - -		tgccttttag	TGGGAAAGCT
	484				
RPS5	TGTGAACCAG	gtgagcctgg	- - 1100[b] - -		gtcttcctag	GCCATCTGGC
	583				
	TGCTGCCAAG	gtgggtgagg	- - 86 - -		ctccttgcag	GGCTCCTCGA
	132				
RPS9	AAGCTGATCG	gtgagtggcc	- - 205 - -		tccccaccag	GCGAGTATGG
	166				
RPS10	GGCCATGCAG	gtaggagggt	- - 232 - -		tgccttacag	TCTCTCAAGT
	368				
RPS11	CCTGCTTCAG	gtgagcgcag	- - 1464 - -		gcctccacag	GGACGTCCAG
	313				
RPS12	CCTAATTAAG	gtaaggctgc	- - 400[b] - -		ctcccaatag	GTTGATGACA
	415				
	AGTAGTTAAG	gtaagtcacc	- - 500[b] - -		ttttttaag	GACTATGGCA
	454				
RPS13	ATTGGAAATA	gtaagtatca	- - 629 - -		gcttttccag	TGAATCATCT
	133				
RPS15A	ATGAAGCATG	gtaagtctgc	- - 750[b] - -		cttttggtag	GTTACATTGG
	378				
RPS19	ACCAAGATGG	gtaagcaggg	- - 480[b] - -		ctctccacag	CGGCCGCAAA
	290				
RPS20	GCCTACCAAG	gtaaagtaaa	- - 423 - -		tttgttacag	ACTTTGAGAA
	152				
RPS21	CGTGGCCGAG	gtgagctggg	- - 168 - -		ctctttctag	GTTGACAAGG
	224				
	TCGTAGGATG	gtgagtgttt	- - 392 - -		ttttcttaag	GGTGAGTCAG
	298				
RPS23	CTTTATTGAG	gtgagtattt	- - 145 - -		tttattccag	GAAAATGATG
	204				
RPS26	GTCTTCGATG	gtaagtgggt	- - 700[b] - -		tattccttag	CCTATGTGCT
	135				
RPS27	AAATGCCCAG	gtgaggagac	- - 344 - -		tttctttcag	GATGCTATAA
	346				
RPS27A	ATATTATAAG	gtgagccagt	- - 465 - -		tgcttttcag	GTGGATGAGA
	241				
RPS28	CTGCTCGCTG	gtgggtgcaa	- - 180 - -		ctgtttacag	GGTCTTGGAT
	192				
RPS29	TTTCATTAAG	gtaggcgtct	- - 2100[b] - -		ttttttttcag	TTGGACTAAA
	1053				
RPL3	CCTCCGCAAG	gtgaggcagg	- - 323 - -		tggcccatag	TCCTTGCTGG
	1173				
	AGCATTCATG	gtgagcacct	- - 206 - -		ctccccacag	GGACCACTGA
	735				
RPL6	AGAAAAAGAG	gtaagtttct	- - 476 - -		ctttatttag	AAATATGAGA
	181				
RPL8	CATCGTCAAG	gtgcggaacg	- - 77 - -		ctgcccacag	GACATCATCC
	323				
	GGCAAGAAGG	gttagcatcg	- - 277 - -		cccctttgcag	CCCAGCTCAA
	478				
RPL9	TCAAATTCAG	gtttgtatgt	- - 214 - -		actcctatag	CGGCTTTGAT
	595				
	GATCTAAGAG	gtaagttctt	- - 307 - -		ttttttacag	TTACCTGGCT

(*Continued on facing page.*)

Implications for the Human Genome Project

The methods we employed for STS generation via intron trapping should be of general use in mapping genes with processed homologs. Recently, PCR-based typing of RH panels (or YAC libraries) was used to map many thousands of gene-based STSs (Schuler et al. 1996). These high-volume gene mapping efforts relied on STSs drawn from 3′ untranslated regions of genes (Berry et al. 1995). This method minimized the chance that PCR products

Table 2. (*Continued*)

Gene	Exon		Intron size (bp)		Exon
	506				
RPL10	CCGCCAGAAG	gtatgtagtg	-- 75 --		tgctccttag ATCCACATCT
	160				
RPL11	TTTTCCAAAG	gtgagtagtc	-- 850[b] --		ttcctgccag CTAGATACAC
	447				
RPL12	GAACTCTCTG	gtaagagcag	-- 347 --		ttacttaaag GAACCATTAA
	528				
RPL13	CGTCCGGAAC	gtaagtgaac	-- 492 --		tctcttctag GTCTATAAGA
	188				
RPL15	GCCAAGCAAG	gtacgtgatc	-- 380[b] --		ttgtgtgtag GTTACGTTAT
	645				
RPL17	GAAGAAAAAG	gtaaataagt	-- 750[b] --		ctttccccag ATATCCCAGA
	347				
RPL18A	ACCCAGTGCT	gtaagctgcc	-- 750[b] --		ctttcaaaag ACCGAGACAT
	457				
	GCAGTTCCAC	gtgagtgccc	-- 143 --		ttcctcacag GACTCCAAGA
	422				
RPL21	AAAGCGCCAG	gtaagaaatt	-- 101 --		cctttaatag CCTGCTCCAC
	365				
RPL23	GAGATGAAAG	gtaggaaatc	-- 147 --		tttaatgcag GTTCTGCCAT
	368				
RPL24	AAGCTATCAG	gtgaggaatg	-- 278 --		taccccacag GGCTGCTAAG
	432				
	TGCTGCTAAG	gtaattatgg	-- 1100[b] --		cttttttacag GCACCTACAA
	379				
RPL27	TTGAAGAGAG	gtaagtaggt	-- 88 --		cccctcctag ATACAAGACA
	334				
RPL27A	GGTGCGATCG	gtaagttaat	-- 800[b] --		gttcttctag GGCTACTACA
	351				
RPL28	CCTGCGCATG	gtgagctggg	-- 195 --		ccgccccccag GCAGCCATCC
	354				
RPL31	CTTTCAAAAA	gtaagttctc	-- 279 --		tcctttacag TCTACAGACA
	85				
RPL34	AAACTAGGCT	gtaagtattt	-- 80 --		cactttctag GTCCCGAACC
	371				
RPL35A	AATCCGAGTG	gtgagtatgg	-- 1700[b] --		ttccctgcag ATGCTGTACC
	145				
RPL37	AAGAGAAAGT	gtaagtaaca	-- 205 --		tatgttttag ATAACTGGAG
	230				
	GCAGATTCAG	gtacagtttg	-- 1400[b] --		catcctgcag GCATGGATTC
	297				
RPL38	CTGCCCCCCG	gtgagtgagc	-- 495 --		ttccctctag GTTTGGCAGT
	70				
RPL41	ATTTTTTTGG	gtgagtgttt	-- 116 --		ttccctgtag AAACCTCTGC
	95				
	GAGAGCCAAG	gtgagcggtt	-- 388 --		cggttgctag TGGAGGAAGA
	118				
	GAATGCGCAG	gtacgttgag	-- 271 --		tgcttttcag GCTGAAGCGC
	869				
RPP0	TGCTGAAAAG	gtaaaaggat	-- 389 --		ctcctttcag GTCAAGGCCT
	394				
RPP1	GCTGCTCCAG	gtaggaaaca	-- 140 --		ttttttctag CTGAGGAGAA

Upper- and lowercase letters denote exon and intron sequences, respectively.
For DDBJ/EMBL/GenBank accession nos., see Table 1.
[a]Last nucleotide in exon is numbered according to position in cDNA sequence.
[b]Size estimated by agarose gel electrophoresis.

would contain introns, thereby reducing the size of PCR products and increasing amplification efficiency. In contrast, our rp gene STSs were designed to span or lie within introns, to identify functional, intron-bearing genes as distinct from processed pseudogenes. Although the procedure for deriving STSs from 3′-untranslated regions of genes is simpler—a requirement of high-volume genomic studies—the procedure would likely yield erroneous mapping results when applied to genes with abundant processed pseudogenes. Had we applied this strategy to rp genes, we would have completely overlooked the functional genes, which are less efficient than their intron-less pseudogene derivatives as templates for PCR amplification with cDNA-based primers. Although an error rate of only 1% has been claimed for the high-volume, STS-based gene mapping efforts (Schuler et al. 1996), these error estimates took no account of processed pseudogenes, the impact of which could be significant if a sizable fraction of all genes give rise to processed pseudogenes. The methods we employed for STS generation via intron trapping should be of use in efforts to systematically map genes with a propensity to generate processed pseudogenes, that is, housekeeping and other genes that are abundantly expressed in the germ line.

Ribosomal Protein Defects in Human Disease

Evolutionary and genetic considerations lead us to predict roles for rp genes in human disease. Ribosomal proteins are highly conserved among eukaryotes and prokaryotes. Virtually all mammalian ribosomal proteins have counterparts (with 40%–88% amino acid identity) in the yeast ribosome (Wool et al. 1996). Of the 78 rat ribosomal proteins whose amino acid sequence is known, at least 49 have recognizable homologs in the archaebacterial ribosome (Wool et al. 1996). Among multicellular animals, the consequences of mutations in rp genes have been explored most thoroughly in *Drosophila.* Here, mutations resulting in reduced expression of individual ribosomal proteins yield the *Minute* phenotype. Because a full complement of ribosomal proteins is required to assemble a functional, stable ribosome, *Minute* cells probably contain fewer ribosomes and thus have less capacity for protein synthesis (Kay and Jacobs-Lorena 1987). Conservation of ribosomal proteins among eukaryotes, combined with sequence studies, indicate that *Drosophila* and human ribosomes are extremely similar. Thus, it is likely

that quantitative deficiencies in human ribosomal proteins, as in *Drosophila,* will result in reduced translational capacity and thereby yield specific, reproducible phenotypes. If specific human phenotypes do result from ribosomal protein deficiencies, those phenotypes may or may not resemble the *Drosophila Minute* phenotype.

The present mapping study was motivated by

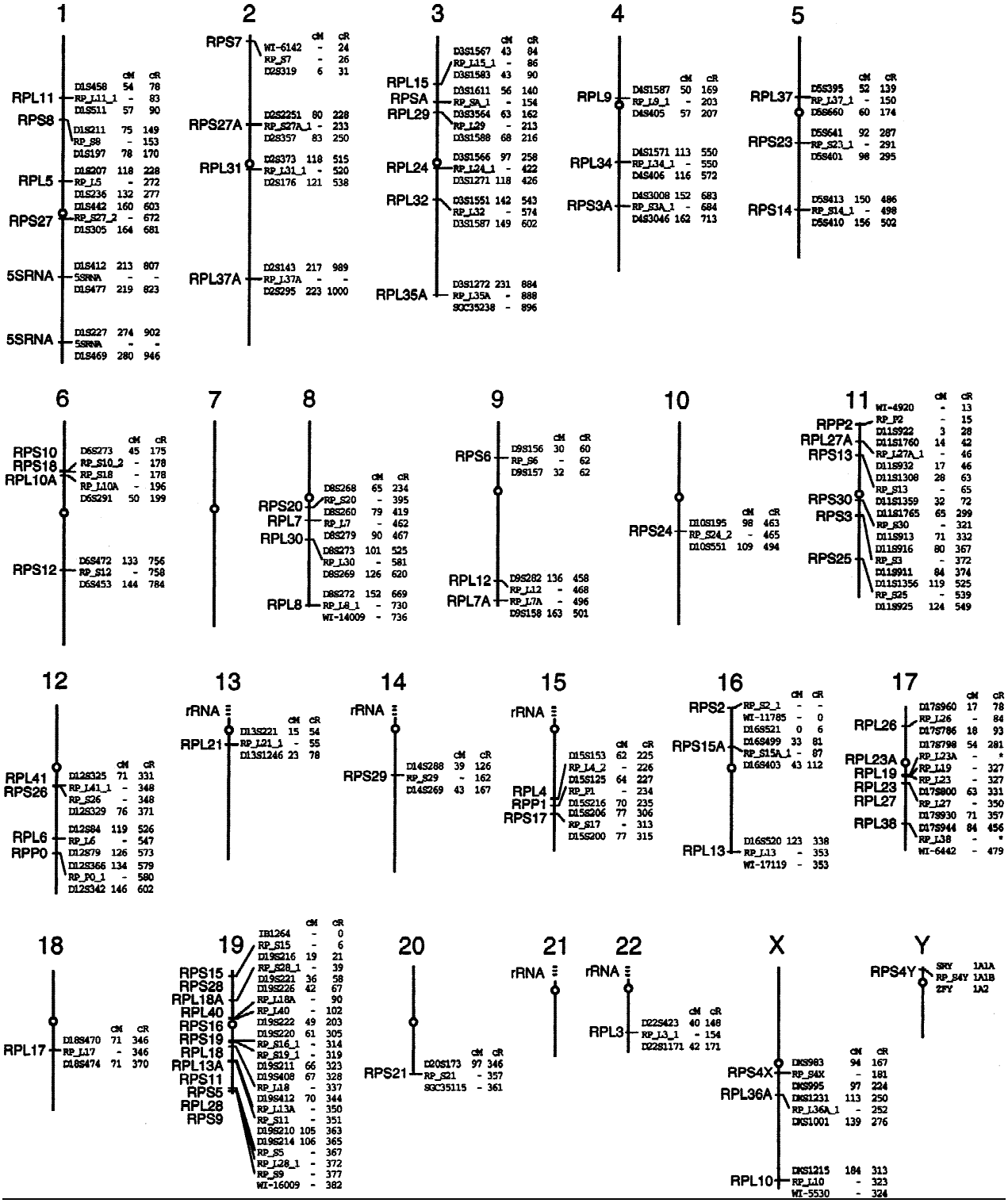

Figure 2 (*See facing page for legend.*)

the possibility that ribosomal protein mutations contribute to human disease, including Turner syndrome, other chromosomal birth defects, and Mendelian disorders. As yet, no human disorder has been traced definitively to a ribosomal protein mutation. Having a map of the rp genes will facilitate the search for mutations and roles in human disease, including monosomies and various Mendelian disorders.

Monosomies

Turner syndrome is a complex human phenotype associated with monosomy X (Ford et al. 1959). Embryos with a 45,X karyotype develop as females with poor viability in utero, and those that survive exhibit short stature, ovarian failure, and specific anatomic abnormalities that may include lymphedema, webbing of the neck, and cubitus valgus (Lippe 1991). These phenotypes likely result from the haploinsufficiency of certain genes that are common to the X and Y chromosomes and that escape X inactivation (Ferguson-Smith 1965); specific XY gene pairs probably account for particular components of this complex phenotype (Zinn et al. 1993) On the basis of studies of individuals carrying partial Y chromosomes, we and our colleagues have previously suggested that ribosomal protein S4, encoded by both the X and Y chromosomes, is an important factor in Turner syndrome (Fisher et al. 1990). Specifically, our results suggest that a quantitative deficiency of RPS4 may be responsible for the lymphedema and neck webbing observed in 45,X individuals, and perhaps also for the poor viability of 45,X fetuses (L. Brown, C. Raut, and D.C. Page, unpubl.). Following the principle, from *Drosophila,* that translational deficiencies stemming from any of a large number of rp genes yield a consistent, reproducible phenotype, we predict that deficiencies of autosomal rp genes may yield Turner-like phenotypes in humans. Given the dispersion of rp genes throughout the human genome, virtually all monosomies and many partial monosomies will entail heterozygous deficiencies of one or more rp genes. We speculate that ribosomal protein deficiencies contribute to the abnormal development and poor viability of monosomic human fetuses and perhaps also to phenotypes observed in children born with partial monosomies. Thus, a search for correlations between ribosomal protein loci and haploinsufficient or haplolethal regions of the human genome should be initiated. [Note: The hypothesized role of RPS4 in Turner syndrome is controversial. Some investigators have concluded that *RPS4* is not a Turner gene, because many Turner patients with structurally abnormal X chromosomes have two or more active copies of *RPS4X* (Just et al. 1992; Geerkens et al. 1996). However, as most such patients do not exhibit the particular Turner features that we attribute to RPS4 deficiency, these observations are consistent with our hypothesis. Omoe and Endo (1996) have suggested that RPS4 is perhaps not involved in Turner syndrome, because some mammalian species that have no *RPS4Y* gene on their Y chromosome exhibit an XO phenotype that is reminiscent of human Turner syndrome. However, because these XO animals do not exhibit the particular Turner features that we attribute to RPS4 deficiency, Omoe and Endo's observations are also consistent with our hypothesis.]

Mendelian Haploinsufficiencies

Quantitative deficiencies of individual ribosomal proteins could possibly result from either gross chromosomal deletions or point mutations in individual rp genes. In the latter case, resulting haploinsufficient traits might display simple Mendelian

Figure 2 A map of genes encoding the human ribosome. The 22 autosomes and two sex chromosomes are shown as vertical lines, on which are positioned 75 rp genes (RP...), five ribosomal RNA gene clusters (rRNA), and two 5S RNA gene clusters (5SRNA). To the *right* of each chromosome are listed rp gene STSs, nearby markers, and approximate distances (in centiMorgans and/or centiRays) from the most distal short-arm marker on maps constructed at Généthon and the Whitehead Institute/MIT Center for Genome Research (Hudson et al. 1995). (Maps shown as of November 1997; for updated maps see http://www-genome.wi.mit.edu/cgi-bin/contig/phys_map). Because of the inherently statistical nature of RH mapping, we have high confidence in marker orders only where markers are separated by at least 15 centiRays. On distal 19q, for example, *RPS5, RPL28,* and *RPS9* appear to be clustered within 10 centiRays and thus cannot be ordered with confidence. Our present data are most consistent with the order RPS5—RPL28—RPS9—qter (as shown), but higher resolution mapping experiments, while confirming the proximity of the three genes, strongly suggest the order RPS9–RPL28–RPS5–qter (N. Kenmochi, G. Lennon, S. Higa, and L. Ashworth, unpubl.). For the Y chromosome, where no genetic map is available, deletion map intervals (Vollrath et al. 1992) are listed. Our assignment of *RPL29* to 3p conflicts with a recent report that it maps to 3q29-qter (Garcia-Barcelo et al. 1997). (*) On chromosome 17, *RPL23A* and *RPL38* were localized to the indicated intervals, but their distances from flanking markers could not be meaningfully estimated.

transmission, probably appearing like autosomal dominant characters in human pedigrees. Because, in *Drosophila,* heterozygotes for loss-of-function mutations in any of a large number of rp genes display the *Minute* phenotype, one might expect the phenotype in human heterozygotes to be similarly consistent despite genetic heterogeneity. We predict that this autosomal dominant phenotype would encompass the components of the 45,X phenotype that are likely attributable to RPS4 deficiency, namely, lymphedema, neck webbing, and, perhaps, reduced fetal viability.

Particularly intriguing is the possibility that Noonan syndrome might be attributable to ribosomal protein deficiency. Many investigators have drawn attention to similarities between the Noonan and Turner phenotypes (Allanson 1987). As one would predict for a *Minute*-like ribosomal protein deficiency, Noonan syndrome is inherited in apparently autosomal dominant fashion, and it is genetically heterogeneous. We note that the *RPL6* and *RPPO* genes map to a region of chromosome 12 implicated, by linkage analysis, in a large pedigree with Noonan syndrome (Jamieson et al. 1994).

Other Mendelian Disorders

In considering the potential range of human disorders that might be caused by rp gene mutations, it is important to recognize that some ribosomal proteins may have additional, extraribosomal functions. Wool (1996) has suggested that, during evolution, proteins of diverse function were recruited to the ribosome to stabilize rRNA or otherwise enhance translation, and that, in some cases, these recruited proteins have also retained their ancestral functions. For example, in mammals and *Drosophila,* RPS3 functions as both a ribosomal protein and an endonuclease (Wilson et al. 1994; Kim et al. 1995; Yacoub et al. 1996). In *Drosophila,* RPS6 functions as a tumor suppressor in the hematopoietic system (Watson et al. 1992), and RPS2 functions in oogenesis (Cramton and Laski 1994), perhaps reflecting extraribosomal functions for these proteins. Given these precedents, human geneticists should not overlook the possibility that ribosomal protein mutations could directly perturb diverse cellular functions, without affecting translation, and thereby produce disease.

METHODS

DNA Sequences and Nomenclature

A complete catalog of rat ribosomal protein cDNA sequences is available (Wool et al. 1996). We used these rat cDNA sequences to query the GenBank, EMBL, and DDBJ databases for human rp gene sequences (cDNA and genomic). The amino acid sequences of homologous rat and human ribosomal proteins are, on average, 99% identical. We refer to the human ribosomal proteins using the established rat nomenclature (Wool et al. 1996).

PCR Primers and Conditions

For most rp genes, the primer selection rules described in the Results proved workable and effective. However, for a few group 3 genes, these rules were too stringent to permit selection of primer pairs, or the primers selected failed to amplify a higher molecular weight product with YAC pools as template. In several such cases, we were able to select satisfactory pairs by picking the forward primer from a larger target region (150–300 bp upstream of the termination codon) or by allowing the forward primer to contain a single AG dinucleotide. In all cases, PCR primer pairs were selected by use of the PRIMER program (S. Lincoln, M.J. Daly, E.S. Lander, Whitehead Institute); optimal oligonucleotide T_m was set at 58°C and the optimal primer length was set at 20 nucleotides.

PCR was performed in 20-µl volumes containing 30–50 ng of template DNA, 10 pmole each of forward and reverse primers, 0.1 mM dNTPs, 10 mM Tris-Cl (pH 8.2), 50 mM KCl, 1.5 mM $MgCl_2$, 5.0 mM NH_4Cl. Reaction mixes were first heated at 90°C and, then, 1 unit of *Taq* DNA polymerase was added. Cycling conditions included an initial denaturation at 94°C for 3 min, followed by 35 cycles of 94°C for 1 min, 61°C for 1 min, 72°C for 1 min, and a final extension at 72°C for 5 min.

Physical Mapping

RP gene STSs were chromosomally assigned by use of National Institute of General Medical Sciences (NIGMS) human-rodent hybrid cell line panels 1 and 2 (Drwinga et al. 1993).

To place rp gene STSs on an existing RH map of the human genome (Hudson et al. 1995), we tested the hybrids of the GeneBridge 4 panel (Walter et al. 1994) in duplicate, by PCR, and analyzed the results using RHMAPPER software (Hudson et al. 1995).

To place RP gene STSs on an existing YAC/STS content map of the human genome (Hudson et al. 1995), we screened 25,344 YACs (plates 709–972) from the CEPH library (Chumakov et al. 1995) using methods described previously (Hudson et al. 1995).

ACKNOWLEDGMENTS

We thank Carl Rosenberg for thoughtful analyses and assistance with primer selection, Lincoln Stein for help in managing the informatics aspects of RH and YAC mapping, Sebastian Gerety for expert assistance in mapping, Helen Skaletsky for help in RH analysis, William Gelbart and Chao Sun for comments on the manuscript, and Ira Wool and Yuen-Ling Chan for encouragement and valuable advice during the course of this work. This study was supported by the Howard Hughes Medical Institute, the National Institutes of Health,

the Ministry of Education, Science, Sports and Culture of Japan, and the Japanese Foundation for Growth Science.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Adams, S.M., N.R. Helps, M.G. Sharp, W.J. Brammar, R.A. Walker, and J.M. Varley. 1992. Isolation and characterization of a novel gene with differential expression in benign and malignant human breast tumours. *Hum. Mol. Genet.* **1:** 91–96.

Allanson, J.E. 1987. Noonan syndrome. *J. Med. Genet.* **24:** 9–13.

Andersson, S., S. Saebøe-Larssen, A. Lambertsson, J. Merriam, and M. Jacobs-Lorena. 1994. A *Drosophila* third chromosome *Minute* locus encodes a ribosomal protein. *Genetics* **137:** 513–520.

Annilo, T., M. Laan, J. Stahl, and A. Metspalu. 1995. The human ribosomal protein S7-encoding gene: Isolation, structure and localization in 2p25. *Gene* **165:** 297–302.

Antoine, M. and M. Fried. 1992. The organization of the intron-containing human S6 ribosomal protein (rpS6) gene and determination of its location at chromosome 9p21. *Hum. Mol. Genet.* **1:** 565–570.

Berry, R., T.J. Stevens, N.A. Walter, A.S. Wilcox, T. Rubano, J.A. Hopkins, J. Weber, R. Goold, M.B. Soares et al. 1995. Gene-based sequence-tagged-sites (STSs) as the basis for a human gene map. *Nature Genet.* **10:** 415–423.

Chumakov, I.M., P. Rigault, I. Le Gall, C. Ballanné-Chantelot, A. Billault, S. Guillou, P. Soularue, G. Guasconi, E. Poullier et al. 1995. A YAC contig map of the human genome. *Nature* **377:** S175–S298.

Colombo, P., M. Read, and M. Fried. 1996. The human L35a ribosomal protein (RPL35A) gene is located at chromosome band 3q29-qter. *Genomics* **32:** 148–150.

Cramton, S.E. and F.A. Laski. 1994. *string of pearls* encodes *Drosophila* ribosomal protein S2, has *Minute*-like characteristics, and is required during oogenesis. *Genetics* **137:** 1039–1048.

Davies, B. and M. Fried. 1993. The structure of the human intron-containing S8 ribosomal protein gene and determination of its chromosomal location at 1p32-p34.1. *Genomics* **15:** 68–75.

———. 1995. The L19 ribosomal protein gene (RPL19): Gene organization, chromosomal mapping, and novel promoter region. *Genomics* **25:** 372–380.

Davies, B., S. Feo, E. Heard, and M. Fried. 1989. A strategy to detect and isolate an intron-containing gene in the presence of multiple processed pseudogenes. *Proc. Natl. Acad. Sci.* **86:** 6691–6695.

Drwinga, H.L., L.H. Toji, C.H. Kim, A.E. Greene, and R.A. Mulivor. 1993. NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2. *Genomics* **16:** 311–314.

Dudov, K.P. and R.P. Perry. 1984. The gene family encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed gene. *Cell* **37:** 457–468.

Fan, W., M. Christensen, E. Eichler, X. Zhang, and G. Lennon. 1997. Cloning, sequencing, gene organization, and localization of the human ribosomal protein RPL23A gene. *Genomics* **46:** 234–239.

Feo, S., B. Davies, and M. Fried. 1992. The mapping of seven intron-containing ribosomal protein genes shows they are unlinked in the human genome. *Genomics* **13:** 201–207.

Ferguson-Smith, M.A. 1965. Karyotype-phenotype correlations in gonadal dysgenesis and their bearing on the pathogenesis of malformations. *J. Med. Genet.* **2:** 142–155.

Fisher, E.M., P. Beer-Romero, L.G. Brown, A. Ridley, J.A. McNeil, J.B. Lawrence, H.F. Willard, F.R. Bieber, and D.C. Page. 1990. Homologous ribosomal protein genes on the human X and Y chromosomes: Escape from X inactivation and possible implications for Turner syndrome. *Cell* **63:** 1205–1218.

FlyBase. 1997. FlyBase: A *Drosophila* database. *Nucleic Acids Res.* **25:** 63–66.

Ford, C.E., O.J. Miller, P.E. Polani, J.C. de Almeida, and J.H. Briggs. 1959. A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). *Lancet* **1:** 711–713.

Garcia-Barcelo, M., P.T.W. Law, S.K.W. Tsui, K.P. Fung, C.Y. Lee, and M.M.Y. Waye. 1997. Mapping of the human ribosomal large subunit protein gene RPL29 to human chromosome 3q29-qter. *Genomics* **46:** 148–151.

Geerkens, C., W. Just, K.R. Held, and W. Vogel. 1996. Ullrich-Turner syndrome is not caused by haploinsufficiency of RPS4X. *Hum. Genet.* **97:** 39–44.

Hariharan, N., D.E. Kelley, and R.P. Perry. 1989. Equipotent mouse ribosomal protein promoters have a similar architecture that includes internal sequence elements. *Genes & Dev.* **3:** 1789–1800.

Hart, K., T. Klein, and M. Wilcox. 1993. A *Minute* encoding a ribosomal protein enhances wing morphogenesis mutants. *Mech. Dev.* **43:** 101–110.

Henderson, A.S., D. Warburton, and K.C. Atwood. 1973. Ribosomal DNA connectives between human acrocentric chromosomes. *Nature* **245:** 95–97.

Hudson, T.J., L.D. Stein, S.S. Gerety, J. Ma, A.B. Castle, J. Silva, D.K. Slonim, R. Baptista, L. Kruglyak et al. 1995. An

STS-based map of the human genome. *Science* **270:** 1945–1954.

Imai, T., K. Sudo, and T. Miwa. 1994. Assignment of the human ribosomal protein S25 gene (RPS25) to chromosome 11q23.3 by sequence analysis of the marker D11S456. *Genomics* **20:** 142–143.

Jamieson, C.R., I. van der Burgt, A.F. Brady, M. van Reen, M.M. Elsawi, F. Hol, S. Jeffery, M.A. Patton, and E. Mariman. 1994. Mapping a gene for Noonan syndrome to the long arm of chromosome 12. *Nature Genet.* **8:** 357–360.

Jones, A.M., R. Marzella, M. Rocchi, and J.E. Hewitt. 1997. Mapping of the human ribosomal small subunit protein gene RPS24 to the chromosome 10q22-q23 boundary. *Genomics* **39:** 121–122.

Just, W., C. Geerkens, K.R. Held, and W. Vogel. 1992. Expression of RPS4X in fibroblasts from patients with structural aberrations of the X chromosome. *Hum. Genet.* **89:** 240–242.

Kas, K., E. Schoenmakers, W. van de Ven, G. Weber, M. Nordenskjold, L. Michiels, J. Merregaert, and C. Larsson. 1993. Assignment of the human FAU gene to a subregion of chromosome 11q13. *Genomics* **17:** 387–392.

Kay, M.A. and M. Jacobs-Lorena. 1987. Developmental genetics of ribosome synthesis in *Drosophila*. *Trends Genet.* **3:** 347–351.

Kim, J., L.S. Chubatsu, A. Admon, J. Stahl, R. Fellous, and S. Linn. 1995. Implication of mammalian ribosomal protein S3 in the processing of DNA damage. *J. Biol. Chem.* **270:** 13620–13629.

Kongsuwan, K., Q. Yu, A. Vincent, M.C. Frisardi, M. Rosbash, J.A. Lengyel, and J. Merriam. 1985. A *Drosophila Minute* gene encodes a ribosomal protein. *Nature* **317:** 555–558.

Kuzumaki, T., T. Tanaka, K. Ishikawa, and K. Ogata. 1987. Rat ribosomal protein L35a multigene family: Molecular structure and characterization of three L35a-related pseudogenes. *Biochim. Biophys. Acta* **909:** 99–106.

Lippe, B. 1991. Turner Syndrome. *Endocrinol. Metab. Clin. of North Am.* **20:** 121–152.

Lomholt, B., S. Frederiksen, J.N. Nielsen, and C. Hallenberg. 1995. Additional assignment of the human 5S rRNA genes to chromosome region 1q31. *Cytogenet. Cell Genet.* **70:** 76–79.

Maeda, N., N. Kenmochi, and T. Tanaka. 1993. The complete nucleotide sequence of chicken ribosomal protein L7a gene and the multiple factor binding sites in its 5′-flanking region. *Biochimie* **75:** 785–790.

Maidak, B.L., G.J. Olsen, N. Larsen, R. Overbeek, M.J. McCaughey, and C.R. Woese. 1997. The RDP (Ribosomal Database Project). *Nucleic Acids Res.* **25:** 109–111.

Mazuruk, K., T.J. Schoen, G.J. Chader, T. Iwata, and I.R. Rodriguez. 1996. Structural organization and chromosomal localization of the human ribosomal protein L9 gene. *Biochim. Biophys. Acta* **1305:** 151–162.

Meyuhas, O., D. Avni, and S. Shama. 1996. Translational control of ribosomal protein mRNAs in eukaryotes. In *Translational control* (ed. J.W.B. Hershey, M.B. Mathews, and N. Sonenberg), pp. 363–388. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Morton, N.E. 1991. Parameters of the human genome. *Proc. Natl. Acad. Sci.* **88:** 7474–7476.

Mount, S.M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* **10:** 459–472.

Nolte, D., G. Taimor, M. Kalff-Suske, and K.H. Seifart. 1996. The human S3a ribosomal protein: Sequence, location and cell-free transcription of the functional gene. *Gene* **169:** 179–185.

Nomura, M., R. Gourse, and G. Baughman. 1984. Regulation of the synthesis of ribosomes and ribosomal components. *Annu. Rev. Biochem.* **53:** 75–117.

Nucifora, G., C.R. Begy, P. Erickson, H.A. Drabkin, and J.D. Rowley. 1993. The 3;21 translocation in myelodysplasia results in a fusion transcript between the AML1 gene and the gene for EAP, a highly conserved protein associated with the Epstein-Barr virus small RNA EBER 1. *Proc. Natl. Acad. Sci.* **90:** 7784–7788.

Oeltjen, J.C., X. Liu, J. Lu, R.C. Allen, D. Muzny, J.W. Belmont, and R.A. Gibbs. 1995. Sixty-nine kilobases of contiguous human genomic sequence containing the alpha-galactosidase A and Bruton's tyrosine kinase loci. *Mamm. Genome* **6:** 334–338.

Omoe, K. and A. Endo. 1996. Relationship between the monosomy X phenotype and Y-linked ribosomal protein S4 (Rps4) in several species of mammals: A molecular evolutionary analysis of Rps4 homologs. *Genomics* **31:** 44–50.

Polakiewicz, R.D., D.J. Munroe, S.N. Sait, K.T. Tycowski, N.J. Nowak, T.B. Shows, D.E. Housman, and D.C. Page. 1995. Mapping of ribosomal protein S3 and internally nested snoRNA U15A gene to human chromosome 11q13.3-q13.5. *Genomics* **25:** 577–580.

Rhoads, D.D., A. Dixit, and D.J. Roufa. 1986. Primary structure of human ribosomal protein S14 and the gene that encodes it. *Mol. Cell. Biol.* **6:** 2774–2783.

Saebøe-Larssen, S. and A. Lambertsson. 1996. A novel *Drosophila Minute* locus encodes ribosomal protein S13. *Genetics* **143:** 877–885.

Schmidt, A., M. Hollmann, and U. Schafer. 1996. A newly identified Minute locus, M(2)32D, encodes the ribosomal protein L9 in *Drosophila melanogaster*. *Mol. & Gen. Genet.* **251:** 381–387.

Schuler, G.D., M.S. Boguski, E.A. Stewart, L.D. Stein, G. Gyapay, K. Rice, R.E. White, P. Rodriguez-Tome, A. Aggarwal et al. 1996. A gene map of the human genome. *Science* **274:** 540–546.

Schultz, J. 1929. The *Minute* reaction in the development of *Drosophila melanogaster*. *Genetics* **14:** 366–419.

Sørensen, P.D., B. Lomholt, S. Frederiksen, and N. Tommerup. 1991. Fine mapping of human 5S rRNA genes to chromosome 1q42.11-q42.13. *Cytogenet. Cell Genet.* **57:** 26–29.

van den Ouweland, A.M., P. Kioschis, M. Verdijk, F. Tamanini, D. Toniolo, A. Poustka, and B.A. van Oost. 1992. Identification and characterization of a new gene in the human Xq28 region. *Hum. Mol. Genet.* **1:** 269–273.

Vollrath, D., S. Foote, A. Hilton, L.G. Brown, P. Beer-Romero, J.S. Bogan, and D.C. Page. 1992. The human Y chromosome: A 43-interval map based on naturally occurring deletions. *Science* **258:** 52–59.

Wagner, M. and R.P. Perry. 1985. Characterization of the multigene family encoding the mouse S16 ribosomal protein: Strategy for distinguishing an expressed gene from its processed pseudogene counterparts by an analysis of total genomic DNA. *Mol. Cell. Biol.* **5:** 3560–3576.

Walter, M., D. Spillett, P. Thomas, J. Weissenbach, and P. Goodfellow. 1994. A method for constructing radiation hybrid maps of whole genomes. *Nature Genet.* **7:** 22–28.

Watanabe, M., A.R. Zinn, D.C. Page, and T. Nishimoto. 1993. Functional equivalence of human X- and Y-encoded isoforms of ribosomal protein S4 consistent with a role in Turner syndrome. *Nature Genet.* **4:** 268–271.

Watson, K.L., K.D. Konrad, D.F. Woods, and P.J. Bryant. 1992. Drosophila homolog of the human S6 ribosomal protein is required for tumor suppression in the hematopoietic system. *Proc. Natl. Acad. Sci.* **89:** 11302–11306.

Webb, G.C., R.T. Baker, M. Coggan, and P.G. Board. 1994. Localization of the human UBA52 ubiquitin fusion gene to chromosome band 19p13.1-p12. *Genomics* **19:** 567–569.

Wilson, D.M.I., W.A. Deutsch, and M.R. Kelley. 1994. *Drosophila* ribosomal protein S3 contains an activity that cleaves DNA at apurinic/apyrimidinic sites. *J. Biol. Chem.* **269:** 25359–25364.

Wool, I.G. 1979. The structure and function of eukaryotic ribosomes. *Annu. Rev. Biochem.* **48:** 719–754.

———. 1996. Extraribosomal functions of ribosomal proteins. *Trends Biochem. Sci.* **21:** 164–165.

Wool, I.G., Y.L. Chan, and A. Glück. 1996. Mammalian ribosomes: The structure and the evolution of the proteins. In *Translational control* (ed. J.W.B. Hershey, M.B. Mathews, and N. Sonenberg), pp. 685–732. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Worton, R.G., J. Sutherland, J.E. Sylvester, H.F. Willard, S. Bodrug, I. Dube, C. Duff, V. Kean, P.N. Ray et al. 1988. Human ribosomal RNA genes: Orientation of the tandem array and conservation of the 5′ end. *Science* **239:** 64–68.

Yacoub, A., L. Augeri, M.R. Kelley, P.W. Doetsch, and W.A. Deutsch. 1996. A *Drosophila* ribosomal protein contains 8-oxoguanine and abasic site DNA repair activities. *EMBO J.* **15:** 2306–2312.

Yon, J., T. Jones, K. Garson, D. Sheer, and M. Fried. 1993. The organization and conservation of the human Surfeit gene cluster and its localization telomeric to the c-abl and can proto-oncogenes at chromosome band 9q34.1. *Hum. Mol. Genet.* **2:** 237–240.

Zinn, A.R., R.K. Alagappan, L.G. Brown, I. Wool, and D.C. Page. 1994. Structure and function of ribosomal protein S4 genes on the human and mouse sex chromosomes. *Mol. Cell. Biol.* **14:** 2485–2492.

Zinn, A.R., D.C. Page, and E.M.C. Fisher. 1993. Turner syndrome: The case of the missing sex chromosome. *Trends Genet.* **9:** 90–93.

Received January 1, 1998; accepted in revised form March 13, 1998.

A Map of 75 Human Ribosomal Protein Genes

Naoya Kenmochi, Tomoko Kawaguchi, Steve Rozen, et al.

Genome Res. 1998 8: 509-523
Access the most recent version at doi:10.1101/gr.8.5.509