

琉球大学学術リポジトリ

Sequence Analysis of a Total of Three Megabases of DNA in Two Regions of Chromosome 8p

メタデータ	言語: 出版者: Oxford University Press 公開日: 2020-06-30 キーワード (Ja): human chromosome 8p22→p21.3 キーワード (En): large-scale DNA sequencing, physical and transcriptional maps, human chromosome 8p11.2 作成者: Isomura, Minoru, Ikegawa, Shiro, Kinjo, Takao, Takeuchi, Kumiko, Yamane-Tanaka, Yuka, Kitami, Kikuko, Nakamura, Yusuke メールアドレス: 所属:
URL	http://hdl.handle.net/20.500.12000/46345

Sequence Analysis of a Total of Three Megabases of DNA in Two Regions of Chromosome 8p

Minoru ISOMURA,^{1,2} Shiro IKEGAWA,² Takao KINJO,² Kumiko TAKEUCHI,² Yuka YAMANE-TANAKA,¹ Kikuko KITAMI,¹ and Yusuke NAKAMURA,^{1,2,*}

Department of Human Genome Analysis, Cancer Chemotherapy Center, Japanese Foundation for Cancer Research, Tokyo, Japan¹ and Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan²

(Received 25 June 1999; revised 29 August 1999)

Abstract

Large-scale sequencing of genomic regions and *in silico* gene trapping together represent a highly efficient and powerful approach for identifying novel genes. We performed megabase-level sequence analyses of two genomic regions on human chromosome 8p (8p11.2 and 8p22→p21.3), after covering those segments with sequence-ready contigs composed of 74 cosmids, 14 BACs, and three PAC clones. We determined continuous nucleotide sequences of 1,856,753 bases on 8p11.2 and 1,210,381 bases on 8p22→p21.3 by combining the shotgun and primer-walking methods. *In silico* gene trapping identified four novel genes in the 8p11.2 region and, in the 8p22→p21.3 region, six known genes (*PRLTS*, *PCM1*, *MTMR7*, *HCAT2*, *HFREP-1* and *PHP*) and three novel genes. The distribution of *Alu* and LINE1 repetitive elements and the densities of predicted exons were different in each region, and *Alu*-rich portions contained more exonic sequences than LINE1-rich areas.

Key words: large-scale DNA sequencing; physical and transcriptional maps; human chromosome 8p11.2; human chromosome 8p22→p21.3

1. Introduction

Large-scale sequencing of genomic regions, and finding genes within those sequences by means of computer software and database searches, is a highly efficient and powerful approach for identifying novel genes.¹ Recent improvements in sequencing technologies have made high-throughput genomic sequencing possible, and computational gene-finder programs such as GRAIL² and GENSCAN³ can identify putative exonic fragments present in anonymous genomic sequences.^{4,5} These data can be integrated via the Internet with partial cDNA sequences that have been archived as expressed sequence tags (ESTs, i.e., exonic fragments), which have been generated with ever-increasing velocity by the Human Genome Project. In addition, whole genomes of several bacterial strains,⁶⁻⁸ of yeast,⁹ and of *C. elegans*¹⁰ have already been sequenced, presenting us with clues for understanding critical features within the human genome through comparative genomics.

Human chromosome 8 appears to be about 135 Mb long and to contain 4000-5000 genes. As part of the Hu-

man Genome Project and to investigate the biological importance of this chromosome, we previously constructed physical maps that included a 10-Mb YAC contig on 8p11.2 (ref. 11) and two cosmid contigs on 8p22→p21.3 that encompassed a region commonly deleted in hepatocellular, colorectal, and non-small cell lung carcinomas.¹² In those experiments we isolated dozens of exon-like sequences by exon-trapping. To complete our understanding on the two loci, and to isolate genes more efficiently from the regions in question, we constructed sequence-ready contigs and performed large-scale sequence analyses. We report here the isolation of several genes from these megabase-level, continuous genomic DNA sequences, by means of *in silico* gene trapping.

2. Materials and Methods

2.1. Construction of sequence-ready contigs

To construct a large-scale sequence-ready contig for 8p11.2, cosmid contigs were constructed from two overlapping CEPH YAC clones (854_f.6 and 937_b.9). Cosmid libraries were constructed from each YAC, and clones containing human DNA inserts were isolated by hybridization selection as described previously.¹³ Contigs were constructed by repeated colony hybridization exper-

Communicated by Toshihisa Takagi

* To whom correspondence should be addressed. Tel. +81-3-5449-5372, Fax. +81-3-5449-5433, E-mail: yusuke@ims.u-tokyo.ac.jp

iments using PCR fragments derived from end-sequences of the cosmids.¹² To fill the gaps in the cosmid contigs derived from the YACs, we screened BAC and PAC libraries (Genome Systems, St. Louis, MO) by PCR according to the manufacturer's protocol, using STSs (sequence-tagged sites) designed from cosmid DNA sequences.

For 8p22→p21.3, cosmid contigs reported previously¹² were connected with BAC and PAC linking clones. The BAC and PAC clones were obtained by means of the Genome Systems PCR-screening system.

2.2. Sequencing of contigs

Nucleotide sequences of cosmid, BAC and PAC clones that comprised minimal tiling paths were determined by combining the shot gun and primer-walking methods. DNAs were prepared from the clones with Qiagen plasmid-purification kits (Qiagen, Hilden, Germany). A 10- μ g aliquot of each DNA was mechanically fragmented by sonication (UD-201 Sonicator; TOMY Co., Tokyo, Japan) and electrophoresed in a 0.8% agarose gel. DNA fractions 2.5- to 6-kb long were recovered from the gel and subcloned into plasmids; pBC was used as a cloning vector for cosmid DNAs, and pBluescriptII for BAC and PAC DNAs. DNAs from plasmid subclones were prepared with an automated plasmid-isolation machine (PI-100, Kurabo, Osaka, Japan). Plasmids containing cosmid-vector sequences were identified by hybridization with vector DNA and eliminated from subsequent analysis. Sizes of the inserts in the subclones were determined by electrophoresis in 0.8% agarose gels; plasmid subclones containing inserts of more than 1 kb were sequenced by a dye-terminator method with T3 and T7 universal primers using an ABI 377 automated sequencer (Applied Biosystems, Foster City, CA). End-sequences of more than five shot gun subclones per 1-kb insert were determined for each cosmid, BAC, or PAC clone. Sequence data from the shot gun clones were assembled by means of the "Phred", "Phrap" and "Consed" programs.¹⁴⁻¹⁶

Any gaps that remained in the assembled sequences were filled with sequences of linking clones by the primer-walking method, using primers synthesized on the basis of assembled sequence. DNA sequencing for primer-walking was performed according to an ABI dye-terminator protocol. Plasmid subclones were used as templates for subclones spanning gaps; otherwise, cosmid, PAC and BAC DNAs served as the templates.

2.3. Computational sequence analysis

The continuous sequences on 8p11.2 and 8p22→p21.3 were examined for the presence of repetitive DNA elements including *Alu*, LINE1 (L1), and THE, using the RepeatMasker program (<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). For computational trapping of genes, repetitive sequences were removed by the RepeatMasker program, and exon prediction was

then performed using the GRAIL (GRAIL2)² and GENSCAN³ programs. For the GRAIL analysis, only "excellent" scores were considered significant.⁴ We used the BLAST algorithm to search for sequence matches against public DNA and EST databases, and FASTA programs against dbEST.

3. Results

3.1. Large-scale, sequence-ready contigs in 8p11.2 and 8p22→p21.3

By assembling cosmid libraries from two YACs on 8p11.2, we obtained four independent cosmid contigs. To fill gaps between these contigs, BAC or PAC libraries were screened using STS sequences corresponding to DNA sequences at the ends of each contig. This strategy allowed us to construct a single 1.9-Mb contig consisting of 42 cosmids, 10 BACs, and 1 PAC (Fig. 1a). In the 8p22→p21.3 region, two gaps in the previously reported contigs¹² (between cosmids 3054 and A266, and between A254 and A014) were filled with BAC and PAC clones. A single contig consisting of 32 cosmids, four BACs, and two PACs, then covered a 1.2-Mb region (Fig. 1b).

3.2. Sequencing and characterization of contigs

On average, the subcloned sequences after removal of vector sequences were ~550 bp long. More than 10 sequences per 1 kb of genomic DNA were determined; hence, the data were considered to be equivalent to more than five-fold redundancy. The sequence of each clone was integrated into a single assembled sequence by comparing overlapping sequences between neighboring clones. In all, we obtained 1,856,753 nucleotides for the 8p11.2 region and 1,210,381 nucleotides for the 8p22→p21.3 region (DDBJ accession No. AP000065-AP000083).

The nucleotide composition of the two regions was similar. The 8p11.2 and 8p22→p21.3 sections contained respectively 29.1% and 30.1% of adenines, 30.1% and 30.4% of thymidines, 20.3% and 19.8% of cytosines, and 20.5% and 19.5% of guanines. In contrast, the two contigs differed with respect to their content of repeat sequences: the 8p11.2 genomic segment contained 572 copies of *Alu* and 290 copies of L1, accounting for 8.4% and 16.6%, respectively, of its entire DNA sequence, while in the 8p22→p21.3 region we identified 491 copies of *Alu* and 147 copies of L1, which accounted for 9.2% (*Alu*) and 10.1% (L1) of the DNA at this locus.

3.3. Computational gene trapping and identification of known genes

To identify expressed sequences within the two genomic fragments of chromosome 8p, we subjected the entire three megabases of sequenced DNA to computational analysis. The GRAIL program predicted 158 tran-

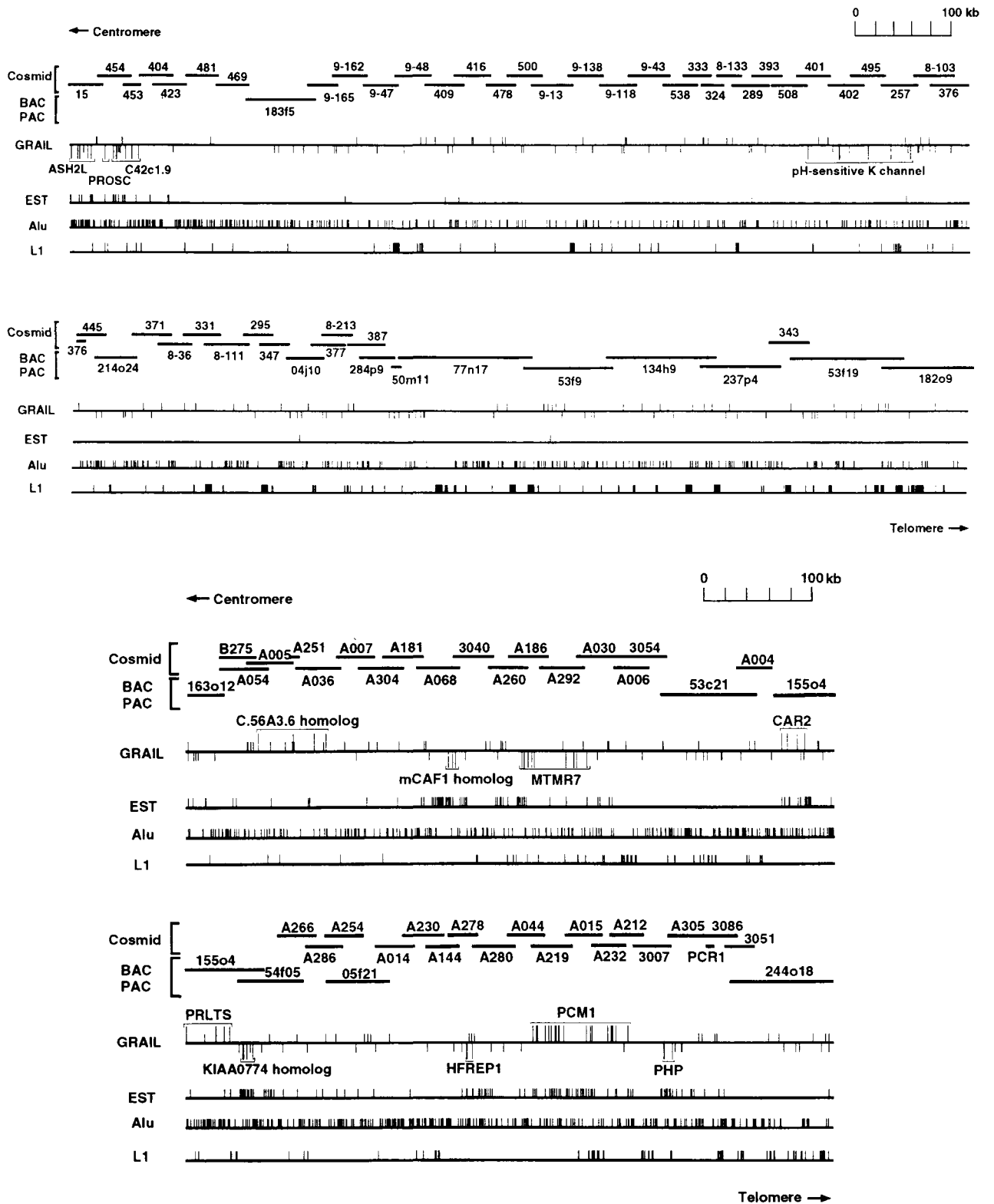


Figure 1. Physical map of a 1.9-Mb region on 8p11.2 (top) and a 1.2-Mb region on 8p21.3 region (bottom). Each figure shows the centromere on the left and the telomere on the right. Cosmid, BAC and PAC clones are shown by transverse lines; locations of exonic fragments predicted by the GRAIL2 program, *Alu* and LINE/1 repetitive elements, and human ESTs are indicated by vertical lines above and below the horizontal lines to indicate presence on the plus and the minus strands (directing centromere to telomere), respectively.

scribed DNA segments from the 8p11.2 region with "excellent" scores. Thirty-nine segments matched archived human ESTs, of which, 19 had also been predicted by the GRAIL program. No previously known genes were present.

In the 8p22→p21.3 region, GRAIL predicted 144 exonic elements, and 196 segments matched sequences in the human EST databases. Sixty segments were identified by both GRAIL analysis and EST search. Six known genes were found within this contig: *PRLTS* (PDGF-receptor beta-like tumor suppressor),¹⁷ *PCM1* (autoantigen pericentriolar material 1),¹⁸ *MTMR7* (myotubularin-related protein 7),¹⁹ *HCA2* (cationic amino acid transporter 2),²⁰ *HFREP-1* (fibrinogen-related protein)²¹ and *PHP* (putative heart protein).²²

3.4. Identification of novel genes

In the 8p11.2 region, six exons (GR61–66; Table 1a) showed significant similarity to the *Drosophila ash2* gene.²³ GR69 showed significant similarity to *C. elegans* hypothetical protein F09E5.8, which is highly homologous to a bacterial proline synthetase-associated gene.²⁴ Eight predicted exons (GR71–78) were highly homologous to a *C. elegans* C42c1.9 gene,²⁵ and ten (GR119–128) to a murine pH-sensitive K⁺ channel gene, *Slo3* (ref. 26).

From the 8p22→p21.3 region (Table 1b), fragments GR230–232 were highly homologous to the murine *CAF-1* (CCR4 associated protein 1) gene.²⁷ Fragments GR280–283 showed a high degree of homology to human KIAA0774 (ref. 28), and GR212–214 to a *C. elegans* C56A gene.²⁵ The nucleotide sequences of these fragments were also identified in the dbEST archive.

4. Discussion

We have determined and characterized the genomic DNA sequence of a 1.9-Mb segment on chromosome 8p11.2 and of a 1.2-Mb segment on chromosome 8p22→p21.3. The GC content of the two regions is almost the same, but the numbers and proportions of repetitive sequences are different. The 1.9-Mb region at 8p11.2 carries fewer *Alu* than the 1.2-Mb region on 8p22→p21.3 (8.4% vs. 9.2% of total nucleotides). In contrast, the 8p11.2 region contains more L1 than the 8p22→p21.3 region does. The inverse proportion of *Alu* and L1 elements has also been suggested by *in situ* hybridization analyses reported elsewhere.²⁹ Both the 8p11.2 and 8p21.3 regions belong to an R-banded region, which has been reported to be relatively rich in *Alu* sequences.²⁹ The 1.9-Mb region at 8p11.2, however, does not have this characteristic, possibly because the fine localization of the 1.9-Mb region is at 8p11.22 which is a G-banded region.

The distribution of repetitive sequences is uneven as

well; in the 1.9-Mb region, the 400 kb at its centromeric end contain more *Alu* than the 400 kb at its telomeric end (1 copy in 2.0 kb vs. 1 copy in 4.0 kb), and the opposite is true for L1 sequences (1 copy in 11 kb vs. 1 copy in 2.5 kb). Interestingly, among the 39 ESTs identified in this portion of 8p11.2, 23 were located within the centromeric *Alu*-rich region. In contrast, *Alu* elements were distributed evenly in the sequenced portion of 8p22→p21.3, and so were the genes.

The two regions also differed in gene annotation. GRAIL analysis identified 144 exonic candidates in the 8p22→p21.3 segment, but only 158 fragments in the 8p11.2 segment (1 exon in every 11.8-kb or 8.4-kb of genomic sequence, respectively). Similarly, the number of ESTs was higher in the 8p22→p21.3 region than in the 8p11.2 region (196 vs. 39).

The distribution of identified genes was also unequal: we identified nine genes from the 1.2-Mb 8p22→p21.3 segment but only four from the 1.9-Mb region on 8p11.2. These observations were similar to reports for other chromosomes, in which for example only three genes were identified from a 685 kb-sequence on 3p21 (ref. 4) but 17 were identified from a 650-kb region on 7q22 (ref. 30), and 20 from a 223-kb region on 12p13.3 (ref. 31). Interestingly, the 8p11.2 segment is rich in L1 and the 8p22→p21.3 segment is rich in *Alu*, in agreement with previous assumptions that *Alu*-rich regions contain more genes than L1-rich regions.^{32,33} Unequal gene density is also reported in the *C. elegans* genome:¹⁰ fewer genes are present in the central parts of worm chromosomes than in the autosomal arms. In keeping with these observations, the gene-poor human 8p11.2 region is adjacent to the centromere.

To date, we have identified numerous genes using the *in silico* method. This approach has several advantages: 1) it can detect expressed sequences independent of their expression levels; 2) it determines the genomic structure of a gene at the same time the gene is identified; and 3) it is not technically demanding. Moreover, this method is very sensitive and reliable. Previously we had screened the 8p22→p21.3 region by exon-trapping but isolated only two genes, *PRLTS* and *PCM1* (ref. 12), trapping only 2 of the 7 exons of *PRLTS* and 3 of the 39 exons of the *PCM1* gene. In contrast, the method used in the present study successfully identified 4 exons of the *PRLTS* gene and 19 exons of *PCM1*.

We have now identified a total of 13 genes in the portions of 8p under study (4 in 8p11.2 and 9 in 8p22→p21.3), including 7 never reported before. The four novel genes in 8p11.2 include orthologues of the *D. melanogaster ash2* gene, the *C. elegans* C42c1.9 gene, a bacterial proline synthetase associated gene, and a murine *Slo3* gene. Detailed characterization of those three *Drosophila*, worm, and bacterial genes have been reported elsewhere.^{34–36} With respect to the fourth orthologue, it is worth noting that murine *Slo3* encodes a pH-

sensitive potassium channel that functions in a voltage-sensitive manner.²⁶ Defects in potassium-channel genes are known to be responsible for neuronal disorders including some kinds of epilepsy.³⁷ Therefore, the human orthologue of *Slo3* might be involved in neuronal performance.

The present study identified six known and three novel genes in the 8p22→p21.3 region. The known genes include *PRLTS*, *PCM1*, *HCA2*, *MTMR7*, *HFREP-1*, and *PHP*. *PRLTS* and *PCM1* had already been isolated and characterized by us.^{17,18} *HCA2* is a cationic amino-acid transporter gene that was mapped to chromosome 8p22→p21.3 by another group.²⁰ *MTMR7* is a myotubularin dual-specificity phosphatase gene, which is conserved from yeast to human, and it has been assigned to 8p22→p21.3 (ref. 19); at least eight human genes belonging to this family of phosphatases have been isolated to date.¹⁹ *HFREP-1* is a fibrinogen-related gene that is often over-expressed in hepatocellular carcinomas.²¹ *PHP* was originally isolated from a heart cDNA library,²² but its function remains unclear. The chromosomal locations of *HFREP-1* and *PHP* were determined for the first time in the present study.

The three novel genes we identified in the 8p22→p21.3 region are orthologues of murine *CAF-1* and *C. elegans* C56A3.6, and a homologue of human KIAA0774. The murine *CAF-1* gene product interacts with CCR4, a protein that is required for regulating a number of genes in yeast.²⁷ *CAF-1* is conserved from *C. elegans* to humans with a high degree of homology, suggesting that it has a critical role in eukaryotic transcription. As the predicted exons designated as GR230–232 showed significant homology to *CAF-1* at the protein level, they may represent another CCR4-related human gene. Characterization of the remaining two genes (KIAA0774 and *C. elegans* C56A3.6 homologues) is in progress. Further study may unveil the significance of these positional candidates for involvement in hepatocellular, colorectal, and non-small cell lung carcinomas.¹²

Acknowledgments: This work was supported partly by the Japan Science and Technology Corporation (JST) and partly by a Grant-in-Aid from the Ministry of Education, Culture, Sports and Science of Japan. We thank H. Hayashi and Y. Hashimoto for technical advice on computational analysis. Details of the sequence-ready contigs and genomic sequences of the two regions are available on the Internet (<http://www-alis.tokyo.jst.go.jp/HGS/>).

References

1. McKusick, V. A. 1997, Genomics: Structural and functional studies of genomes, *Genomics*, **45**, 244–249.
2. Uberbacher, E. C. and Mural, R. J. 1991, Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach, *Proc. Natl. Acad. Sci. USA*, **88**, 11261–11265.
3. Burge, C. and Karlin, S. 1997, Prediction of complete gene structures in human genomic DNA, *J. Mol. Biol.*, **268**, 78–94.
4. Ishikawa, S., Kai, M., Tamari, M. et al. 1997, Sequence analysis of a 685-kb genomic region on chromosome 3p22-p21.3 that is homozygously deleted in a lung carcinoma cell line, *DNA Res.*, **4**, 35–43.
5. Claverie, J. M. 1997, Computational methods for the identification of genes in vertebrate genomic sequences, *Hum. Mol. Genet.*, **6**, 1735–1744.
6. Fleischmann, R. D., Adams, M. D., White, O. et al. 1995, Whole-genome random sequencing and assembly of *Haemophilus influenzae*, Rd., *Science*, **269**, 496–512.
7. Fraser, C. M., Gocayne, J. D., White, O. et al. 1995, The minimal gene complement of *Mycoplasma genitalium*, *Science*, **270**, 397–403.
8. Blattner, F. R., Plunkett, G. 3rd, Bloch, C. A. et al. 1997, The complete genome sequence of *Escherichia coli* K-12, *Science*, **277**, 1453–1474.
9. Goffeau, A., Aert, R., Agostini-Carbone, M. L. et al. 1997, The yeast genome directory, *Nature*, **387** (supp.), 5–105.
10. The *C. elegans* sequencing consortium, 1998, Genome sequence of the nematode *C. elegans*: a platform for investigating biology, *Science*, **282**, 2012–2017.
11. Koyama, K., Sudo, K., and Nakamura, Y. 1995, Isolation of 115 human chromosome 8-specific expressed-sequence tags by exon amplification, *Genomics*, **26**, 245–253.
12. Chinen, K., Isomura, M., Izawa, K. et al. 1996, Isolation of 45 exon-like fragments from 8p22→p21.3, a region that is commonly deleted in hepatocellular, colorectal, and non-small cell lung carcinomas, *Cytogenet. Cell Genet.*, **75**, 190–196.
13. Tokino, T., Takahashi, E., Mori, M. et al. 1991, Isolation and mapping of 62 new RFLP markers on human chromosome 11, *Am. J. Hum. Genet.*, **48**, 258–268.
14. Ewing, B., Hiller, L., Wendl, M. C., and Green, P. 1998, Base-calling of automated sequencer traces using Phred. I. Accuracy assessment, *Genome Res.*, **8**, 175–185.
15. Ewing, B. and Green, P. 1998, Base-calling of automated sequencer traces using Phred. II. Error probabilities, *Genome Res.*, **8**, 186–194.
16. Gordon, D., Abajian, C., and Green, P. 1998, Consed: a graphical tool for sequence finishing, *Genome Res.*, **8**, 195–202.
17. Fujiwara, Y., Ohata, H., Kuroki, T. et al. 1995, Isolation of a candidate tumor suppressor gene on chromosome 8p21.3-p22 that is homologous to an extracellular domain of the PDGF receptor beta gene, *Oncogene*, **10**, 891–895.
18. Ohata, H., Fujiwara, Y., Koyama, K., and Nakamura, Y. 1994, Mapping of the human autoantigen pericentriolar material 1 (*PCM1*) gene to chromosome 8p21.3-p22, *Genomics*, **24**, 404–406.
19. Laporte, J., Blondeau, F., Buj-Bello, A. et al. 1998, Characterization of the myotubularin dual specificity phosphatase gene family from yeast to human, *Hum. Mol. Genet.*, **7**, 1703–1712.
20. Hoshida, R., Ikeda, Y., Karashima, et al. 1996, Molecular cloning, tissue distribution, and chromosomal localization of human cationic amino acid transporter 2

- (HCAT2), *Genomics*, **38**, 174–178.
21. Yamamoto, T., Gotoh, M., Sasaki, H., et al. 1993, Molecular cloning and initial characterization of a novel fibrinogen-related gene, HFREP-1, *Biochem. Biophys. Res. Commun.*, **193**, 681–187.
 22. Churchill, J. R., Hoffman, S., and Wieland, S. J. 1995, A new gene family predicted by a novel human heart cDNA, *Mol. Biol. Cell*, **6S**, 418a.
 23. Adamson, A. L. and Shearn, A. 1996, Molecular genetic analysis of *Drosophila ash2*, a member of the trithorax group required for imaginal disc pattern formation, *Genetics*, **144**, 621–33.
 24. Savioz, A., Jeenes, D. J., Kocher, H. P., and Haas, D. 1990, Comparison of proC and other housekeeping genes of *Pseudomonas aeruginosa* with their counterparts in *Escherichia coli*, *Gene*, **86**, 107–111.
 25. Wilson, R., Ainscough, R., Anderson, K. et al. 1994, 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*, *Nature*, **368**, 32–38.
 26. Schreiber, M., Wei, A., Yuan, A. et al. 1998, Slo3, a novel pH-sensitive K⁺ channel from mammalian spermatozoa, *J. Biol. Chem.*, **273**, 3509–1356.
 27. Draper, M., P., Salvatore, C., and Denis, C., 1995, Identification of a mouse protein whose homolog in *Saccharomyces cerevisiae* is a component of the CCR4 transcriptional regulatory complex, *Mol. Cell. Biol.*, **15**, 3487–3495.
 28. Nagase, T., Ishikawa, K., Suyama, M. et al. 1998, Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro, *DNA Res.*, **5**, 277–286.
 29. Korenberg, J. R. and Rykowski, M. C. 1988, Human genome organization: *Alu*, lines, and the molecular structure of metaphase chromosome bands, *Cell*, **53**, 391–400.
 30. Golckner, G., Scherer, S., Schattevoy, R. et al. 1998, Large-scale sequencing of two regions in human chromosome 7q22: analysis of 650 kb of genomic sequence around the EPO and CUTL1 loci reveals 17 genes, *Genome Res.*, **8**, 1060–1073.
 31. Ansari-Lari, M. A., Shen, Y., Muzny, D. M. et al. 1997, Large-scale sequencing in human chromosome 12p13: experimental and computational gene structure determination, *Genome Res.*, **7**, 268–280.
 32. Soriano, P., Meunier-Rotival, M., and Bernardi, G. 1983, The distribution of interspersed repeats is nonuniform and conserved in the mouse and human genomes, *Proc. Natl. Acad. Sci. USA*, **80**, 1816–1820.
 33. Bernardi, G., Olofsson, B., Filipinski, J. et al. 1985, The mosaic genome of warm-blooded vertebrates, *Science*, **228**, 953–958.
 34. Ikegawa, S., Isomura, M., Koshizuka, Y., and Nakamura, Y. 1999, Cloning and characterization of ASH2L and Ash2l, human and mouse homologs of the *Drosophila ash2* gene, *Cytogenet. Cell Genet.*, **84**, 167–172.
 35. Ikegawa, S., Isomura, M., Koshizuka, Y., and Nakamura, Y. 1999, Cloning and characterization of a novel gene (C8orf2), a human representative of a novel gene family with homology to *C. elegans* C42.C1.9, *Cytogenet. Cell Genet.*, **85**, 227–231.
 36. Ikegawa, S., Isomura, M., Koshizuka, Y., and Nakamura, Y. 1999, Cloning and characterization of the human and mouse PROSC (proline synthetase co-transcribed) genes, *J. Hum. Genet.*, **44**, 337–342.
 37. Biervert, C., Schroeder, B., C., Kubisch, C. et al. 1998, A potassium channel mutation in neonatal human epilepsy, *Science*, **279**, 403–406.

Table 1. Summary of predicted exonic sequences. A "+" indicates prediction by the GENSCAN program, matching with ESTs, or isolation by the exon-trap method. "Gene name" indicates the archived gene product for which FASTA analysis detected the highest degree of similarity to the peptide predicted from each putative exonic sequence. The overlap column indicates the lengths of homologous sequences. Tables 1a and 1b are for the 8p11.2 and 8p22→p21.3 regions, respectively.

exon name	EST	size (aa)	Predicted Amino Acid Sequence	Gene Name	identity (%)	overlap a.a.
GR 1		58	MSKRILFNDLHCLGVSSAQRSLPLPSPQSSSSGLINKDDNNHHIALVVLKNI	hypothetical protein [Arabidopsis]	37.2	59
GR 2		35	KNEDEVESVQSSNSVIMDGGVETGFSTCVHVM	PERB11.1 protein [Homo sapiens]	52.6	19
GR 3		43	TVSKLSNNIKYLHMSSTHTNGQWFTKLVEDTVLPREMFVL	T21D12.9a gene product [C.elegans]	36.5	41
GR 4		63	IFSDHDGIKLEINNKNFNNTWTKLNNILLNDQWVKEEIKKEIKFLGTN	HYPOTHEICAL 16.3 KD PROTEIN C02F5.10	80.0	10
			NNRKNIPKPM			
GR 5		39	DAVVMDDPKARHPAKPKDVENLYCSQIHGPHFTVITVQ	ALP-246 gene product	36.3	33
GR 6		54	DEHSMKNHQTSESHQNEINETHKFLNTSEQRGLGFCRLELETLTILMLTSEQ	(-)		
GR 7		23	MHVNGGWDGYMRFLDSGMVMV	lipase [Pseudomonas sp.]	50.0	22
GR 8		39	FEVFNAYFPLVSRLEIPQFFICPRMLEIGNIQDTAVNT	antigen - Plasmodium vivax	47.8	23
GR 9		14	GEEDWEDQIIIGFF	HYPOTHEICAL 28.3 KD PROTEIN IN QOXD-VPR	63.6	11
GR 10		45	KDITDGYRGKEDVKTQGEYGLYKQORDASEETNPVDTLIDFFQPE	(-)		
GR 11		37	SHRNAKKECYLVLWEHLGLDHLHKEVREVFPEEMIE	fibrinogen B beta - Rattus norvegicus	35.1	37
GR 12		70	KEEGTKLEGENVAFVQSQKGVVPRNRTFQVSLPLGSPKPEEGGQEKRQ			
			TWGSDFSERLHEEVMEKP			
GR 13		38	ISDEVLQKIGGIKPHFVKRRSGNNWVSIFFQSFQKWQH	G-protein subunit [Arabidopsis thaliana]	60.0	15
GR 14		30	MAQSLGAAGTPLESTFIPHVAGAAPDKDSK	hypothetical protein Rv3613c	60.0	20
GR 15		62	TOISLHLVPGDQGNKDFSSHGAASSISLVLPAGDEGAGHLFDCRIIYV	type II collagen [Homo sapiens]	42.4	33
			AVTSNAIIFV			
GR 16		17	MDVVMKAIESFLKIFIK	INTERLEUKIN-6 PRECURSOR	35.2	17
GR 17		12	AGSYDFSKCQK	calgranulin C [Oryctolagus cuniculus]	60.0	10
GR 18		33	YVLCVFTLTLTLLSKTVMDLHNIILTDKETEAOQN	Na/Ca,K-exchanger [Caenorhabditis]	50.0	18
GR 19		31	FDKVDELIIFVISQSSRNIVVGEFLQVDGLG	interphotoreceptor retinoid binding	45.0	20
GR 20		13	MGRQYDEGKCGE	crotoxin A1:SUBUNIT=beta1:ISOTYPE=1 -	72.7	11
GR 21		123	LIAPIEKVTVVWIEDQTSNMSLRQSLIQRKALTFNFKAERGKAESEKLL	probable transposase - human	77.1	118
			EATRGWFMIFKIRSQLHYKVGESTSADVEAVVSPDDLTKINDEGFLTKQ	transposase		
			QIFSVDATAFYWRNDVHS			
GR 22		44	RKEEKREKEREDERNGAKLTSTPVIILVLDVAANENQSMWL	C42D4.6 gene product	32.6	49
GR 23		56	TEANPELDESVPQRHREGSGIRELVEFILLSVEMMPNGGLQLVGMGEN	FLORICAULA PROTEIN.	45.7	35
GR 24		44	NTSDHVMLEVMESRADTSLSDPAGAWKSCFSQEKAYPQDSVQKL	hypothetical protein - Pseudomonas	36.3	44
GR 25		33	PHFFTKVQVTCNSHINMLTSSQDPNKNLDKLA	endo-1,4-beta-glucanase - Pyrococcus	40.0	30
GR 26		52	STINVIDSHYHEKPCNANANLNFCKHAKWDSQGRQLHGPRIEYLLIQIRN	putative second envelope polyprotein	35.7	28
GR 27		38	YGDPTMLKGDYKDTCKNAEINRNEHEATLIGKNVT	copN gene product [Chlamydia psittaci]	38.2	34
GR 28		44	MDQVHTIQAASITKSLCPSLTSHEFLALKGMITDSLDKAWMMPD	membrane-type 1 metalloproteinase	41.9	31
GR 29		39	ENVVTLATVDFISNATALFPWMSDLKYPDWLDRKIISY	fork head domain protein [Xenopus]	42.8	28
GR 30		66	MRRSGPKRKAVAEANITEAEVDQMSLILLLLFSYVNSVSGEHRKHLCSAQPNSPD	PROBABLE CONJUGAL TRANSFER PROTEIN	35.7	42
			FLQFFKTFDGAEREK			
GR 31		52	VCCCRTDELPKTLYPPHKAVSSENFNSGKCECIDIIHSIPNADENLPLK	HYPOTHEICAL PROTEIN MJ1597.1	31.4	35
GR 32		20	VNYVMEALLMVVHYCLFFPK	cytochrome c oxidase subunit 1	58.3	12
GR 33		40	KEFKGELVVEWKKTALLKQCCYSSVTAPAPRACPIGSELR	Synechococcus PCC7942 nucleoside	40.0	30
GR 34		36	MRKSLGLNDLGHQVITMERVVTGCGCHGNGKLAWHT	heat shock protein [Arabidopsis]	40.6	32
GR 35		36	MFLMNSRQDKEDREEGKEENEWKCGSVNRCKTSLV	PA28 alpha subunit [Homo sapiens]	33.3	30
GR 36		25	NNHPDKSTAINIEARPSISKMITDY	trichotecene 3-O-acetyltransferase	38.4	26
GR 37		57	TMINCAGNTEAKNLYVISYKGIKGBELNRLPAAAGMDMMMATVKKGTPELRKK	ribosomal protein L23	82.1	56

GR 38	255	MGSSRNILGSGRWVPTSMELKVLKQLEAQSTEKERAFADRYGWAFLTVPQEVHT QSLRDAAWRALQTLVGHLEAQHNSEKQLEAINGELQAQAGHLKQLQSL EKELEAAVNAAGLGPSSQPEPLSLIKLEKQVILYGHSHYFNFPVMPVRKPDG TWQMTVDYWEINLKVTPPLHAAVLSIMELMDRLVVELAQYHYVVDLANAFFSI DITPESEQEAFASMDGSGQWTFVLPQDYVHSPTIYHGLVDDVMLTFD	polyprotein	39.0	164
GR 39	28	KFKPLPALKMEEECPDSRNVVASSIMO	trefoil factor pS2 precursor	53.3	15
GR 40	36	SSEEVNNEHKMLSENMICITLHIKIKITNTVTDI	hypothetical protein [Escherichia coli]	43.3	30
GR 41	51	IGTISGRMATSLSLGNKYRREGGLNNMHIFCTVLEPGKSKIKILADLVL	variable region of immunoglobulin lambda	28.2	46
GR 42	17	EKDAMDQVAEKQKQRN	C18C4.5 gene product	52.9	17
GR 43	29	DDOKGFLYGVYQFQSHKTVTAVHPVTTG	40S ribosomal protein	42.8	28
GR 44	51	MSHIDAKKIACVIYIMPKCFSEHFRKLVPSFEMERNPKSTKEVYVDLVNMEV	ZEAXANTHIN GLUCOSYL TRANSFERASE	29.7	37
GR 45	56	MAQACSLCCGDHARCAQGERHEHLILISMVHCKKDSQIQLKAKKIYIVRIT	(-)		
GR 46	30	DDNSEKEQHICFELAHRSYLLWFSVFA	Antennal binding protein X	33.3	24
GR 47	34	TWPKCHEREDESKRNTVILITTKETMRSHFLKQG	desmoglein 3 [Mus musculus]	60.0	20
GR 48	30	MDSKFPQNNLSNICYGDLTYRDIYKEASA	putative protein [Arabidopsis thaliana]	34.4	29
GR 49	32	MLNVIDTSMINALISQLFNWVSELCARYSA	S-ADENOSYLMETHIONINE SYNTHETASE	52.6	19
GR 50	47	GSVIPNAVSVFAIFGNASVSELMSCFLWALEETPIPEHLTPAQQA	TRYPsin/FACTOR XIIA INHIBITOR PRECURSOR	37.5	32
GR 51	49	GFQELNVIVDDHFEVQVETETCFENRQPAASCLQCKWSTRLNILEEVL	STP1 gene - Saccharomyces cerevisiae	33.3	36
GR 52	56	APGPEHPLLAENCYDLSVLLRGRLPTRVWLIHDDGCGIWAIMSWRIVVY	DIAMINOPIMELATE EPIMERASE	37.2	43
GR 53	33	RHLVWRPQNIIMEPQSEWAKVHLVRIHLKIK	icb-1 gene	38.4	26
GR 54	33	HKTTTFEKVPHIQEGISAQSVQEEESRQSL	ORF 172 - Pyrococcus horikoshii	29.0	31
GR 55	34	NLSASRNKRKCFQSEGLDLTINTLGGKDGLE	F57H12.6 gene product	39.2	28
GR 56	18	NWVVLGITDFKNEADPH	(-)		
GR 57	41	VKLQTFVTSVTKLVARLELFPVPPGVLVLLGSGWKLQIFA	(-)		
GR 58	38	NSKGSNDEGSAADIPKSIVRSQVIGISWTVQVILLFHH	envelope glycoprotein	38.7	31
GR 59	71	VHEENMTKTFEAFKSLYLEVINVAISHIVLAKASHVTFNKVKEKHNLIKKKR	argininosuccinate lyase	27.6	76
GR 60	19	QOKEGWLNGANDTATYPIV	putative secreted solute binding protein	57.1	14
GR 61	55	PQLKISDDRLLTVGKGYSMVRASHGVRKGAWEYFEITVDEMPDPTAARLGMS	ash212	100	55
GR 62	65	RLPPHGYLPHFPNKDGYRIILAEPPHAPDPEKLELDCWAGKPIPGDLYRA	ash212	100	65
GR 63	30	CLYERVLLALHDDR	ash212	100	30
GR 64	34	GIAAGSSGKGRGAKRKOQDGGTTGTTKKAR	ash212	100	32
GR 65	36	SKERDVLVKEHPDPGSKDPEEDYPKFGLLDQVH	ash212	100	31
GR 66	49	LKEMCLALANLWQSRITQDEHPKTFMSKDKVEVEL	ash212	100	22
GR 67	22	EGAGDTSEVMDTQAGSVDEENGRQLGEVLELQCGICTKWFTADTFGIDTS	(-)		
GR 68	12	EANLVDVYGGLETESNKGKDTL	(-)		
GR 69	36	VOELLEKASNPK	30S RIBOSOMAL PROTEIN	72.7	11
GR 70	65	DLPAIQRLVAVSKTKPADMVI EAYGHGQRTFGENY	F09E5.8 gene product	61.2	31
GR 71	57	MLIHCYSGMILSNRYLYLKSYNSESKMVILTFIYLSLMEINYLASGYC	hypothetical protein	41.3	29
GR 72	26	VCECNQYFFKLYS	C42c1.9	77.7	27
GR 73	47	LMKYKATASNSKIYFGKDI PNMEMDSAGSVSKQFEGLDKLSFGLEDEPLET	C42c1.9	76.9	26
GR 74	36	AAFLAREKAKADAECYTKAMKIAEANK	C42c1.9	79.3	29
GR 75	46	AEKVAOAEIITYGQKVMKEKTEKIKISELEGKQKQSCLSPPPTPCG	C42c1.9	93.3	30
GR 76	24	ESEKTLIIAAQKQKVEKEAETEKKALIGLNVVL	C42c1.9	78.2	23
GR 77	20	KIKAOEWELGSLTTLTFPSAAVLISLQAVRVTKPIPEAIRRNYEL	C42c1.9	79.1	24
GR 78	40	QIDENLKLALQDQLTSMAPGLVIQ	C42c1.9	80.0	20
GR 79	40	GGVMYFDRIEVVNFVPA	C42c1.9	42.8	35
GR 80	57	DKGSLMAQLGAVVAVASSFFCASLFSVHKIEEGHIGVY	(-)		
GR 81	33	VISRRRSARQREPAILEQYAPSDSPDNRSDDGSGLEKWCMEPGSDTSAPPLA	B lymphocyte activation antigen B7-1	47.6	21
GR 82	28	TTAFIVIIIIIIISKCFMRHLHCASIWRRLLIDA	type I restriction-modification enzyme 2	50.0	22
GR 83	28	LVNGRAEVGLLAYLTVNSGFSQNFCTPR			

GR 82	EAIPLAPHDIEKKKFLTCVQDGG	ORNIITHINE DECARBOXYLASE	47.3	19
GR 83	MEEEEEEEEREREEMQEEKKEEEEEEEGEMEEKEXEKEXEKEXEAAAAEAGW RWRGGCCNNGKAGSCL	cyclic nucleotide-gated channel beta subunit	66.0	50
GR 84	MEEEEEEEEEEEEEEEEEEEEEEEEEEEEEKKEKE	GLUTAMIC ACID-RICH PROTEIN PRECURSOR	69.2	39
GR 85	HSEQEAEEVEEEEEEEKKEEEEEIEERKEEEMEEEEEGERE	troponin T - fruit fly	71.7	46
GR 86	VEVMSERDHERFKYGYGIVSITNSLKIKESGNKLA	LYSYL-TRNA SYNTHETASE	39.3	33
GR 87	DLKQGDGSKENEGTNQLNWMTQWLTKKAEEEL	T19F06.5 gene product	42.8	28
GR 88	RLLRKEQFVPKIP	thrombospondin-related protein	50.0	14
GR 89	VLVLKGISLQVGLASWIVSWLRTEATGYTFALLPPGTHHTEQPPSKHEQNGA ELFCNCVSCFEDPC	maif gene	43.4	46
GR 90	MFSPTCPSTVCHMRMQEPRHQTDAGVMILDFPDSGT	hypothetical protein slr0981	33.3	36
GR 91	LEKSENIIISM	OLIGOMYCIN RESISTANCE ATP-DEPENDENT	63.6	11
GR 92	ISIIIIIIIIIIIIIAIIVPNSNEYITISLQHAPATLLVIVSHNYLGLILDS	hypothetical protein YM9958.10	35.1	37
GR 93	KPGDHGREILPAALLSSNLGRILFLLTPRGLLSCSGTSPAFTKDA	(-)		
GR 94	VTEHVDSSQSKKKI	Rat NKAAL gene for Na+/K+ -ATPase alpha	63.6	11
GR 95	MSTRESTQKYGHRMYAQHCTRPWRGKVSALKVEVTVOEPDAGYPPQVQ	P16 protein - Mycoplasma hypopneumonia	33.3	36
GR 96	LVHKDAEGLMQKDQICRYKISDKEDYQKHEKEPESNMVAVRRLCMG	T. pallidum predicted coding region	39.4	38
GR 97	ETDVTHGEKEQYGTAAHLRAWTSKGAYSPOPREV	APURINIC ENDONUCLEASE-REDOX PROTEIN	35.4	31
GR 98	SAYLLDQSLYRVEGETEARNSDONNAAEQGFEFLTKPTIEA	HYPOTHETICAL PROTEIN HI1594	38.4	39
GR 99	MINTDNAKFVYHFEVILVFNKCTP	NITRATE REDUCTASE	50.0	20
GR 100	MDTGRKSGSRVTMNAHCQHVRMVQ	IS600 gene product [Escherichia coli]	40.0	25
GR 101	GKEEVLQYHKQAIQGGVQVFVFELELPDKE	calreticulin precursor	41.1	34
GR 102	KEHEIYKVEDTEGEQRPRKKEKENEKERSGREERLDGRKDKD	pp52/S37 gene - Mus musculus	38.6	44
GR 103	MSIAENGINEETEKKLTYTLNTERKEKEEREITDN	flagellar filament cap protein	42.4	33
GR 104	EKTDQSNRVNMSGSTA	protein, large T - monkey B-lymphotrop	50.0	16
GR 105	VMKLIPVLMDDIERFKASTEVEVTTDVVETARELEVEPNS	Human Tigger1 transposable element,	78.3	37
GR 106	SKGEEKDDKNIPMEMEAREFEEMTESQONSEEGTFFSDKESQEWVDSMA EEGSDSNIIGSDSNTIVQEPKPIPEDEKRR	BAF57 protein	79.7	84
GR 107	NESSDLVANGSFSVNP	probable membrane protein Ylr444c	60.0	15
GR 108	EEMEARVQ	core-binding factor beta subunit p17.6	77.7	9
GR 109	SGADAEVLGLFDSFNTRDLLNTMCKQKKYLLRWQKIFITSKVVVIA	replication protein E1	53.3	30
GR 110	TFYDVKIFHVCADDYGGSWPLVAEAHLK	immunoglobulin heavy chain variable	47.0	17
GR 111	VEQLGIFGDTSFKRWYRQQGTPAQV	HYPOTHETICAL 19.2 KD PROTEIN IN ASN2-	52.1	23
GR 112	RNSRSLNYYTKGIPKAVAIIVGIDPDLT	PUTATIVE 60S MITOCHONDRIAL RIBOSOMAL	48.0	25
GR 113	SNEVSKTLKGNPSSKKNFE	cro protein [Bacteriophage H-19B]	41.1	17
GR 114	MLTLNKLGIKIKITRAIYDKHTTITNLNGQKLEEFTLRTEMRQGCPVS PLLFNIIILEVLARAI	ORE III T beta G41	80.5	67
GR 115	EKRKRDRDADLAL	fusion protein from strain BTA 1686	57.1	14
GR 116	PRNTFQGLFCGSLDLFLGLVGLYRIIDEEELNPENKRGVC	pH-sensitive K channel - Mus musculus	80.0	40
GR 117	SYONGNKKDKHNMSHHTHQKLLKEARMKTIITGIRVSELIILLITYPPWQ CHHKDIKEMAFQASYPPRIKQITL	T18E12.15 gene product	28.3	60
GR 118	IKKLELKRSDLPETLISVRAQTLQOQFDPGPAYSPYT	(-)		
GR 119	AFYNYHVLELLQMLVTVGGVSSQLEHLDKDKVYGVADSCITLLSGRNRLKLG LLSLHETILSDVN	pH-sensitive K channel - Mus musculus	61.5	65
GR 120	NPSNIHFIEQLGGLGSLQETNHLSTAFSTGTVFSGFLDILLAT	pH-sensitive K channel - Mus musculus	78.2	46
GR 121	GCALYSGDLHAANIEQCSMCVAVLSPPPQSSNQTLVDTTEAIMALTIGSLQI DSSSDFPSVSVS	pH-sensitive K channel - Mus musculus	60.3	63
GR 122	TLQHDVEQDSDQDSSGMFHWCKPRTSLDKVTL	pH-sensitive K channel - Mus musculus	62.0	29
GR 123	ALFYCSVCHDDVFIPELITNCGGKRSRQSHIT	pH-sensitive K channel - Mus musculus	59.3	32
GR 124	MPKQTWKHKHFLNSMKNKILTQLRSDDFAGMSFPEVAR	pH-sensitive K channel - Mus musculus	88.8	36
-GR 125	FIVCGNITVDSVTAFLRNFLNRDKSGEINTEIVCLG	pH-sensitive K channel - Mus musculus	88.8	36

GR 126	30	ILFANYIPEMVELFANKRKYTSSEYALKGK	pH-sensitive K channel - Mus musculus	76.6	30
GR 127	24	LRFLRALRLLELPQILQLRAIKT	pH-sensitive K channel - Mus musculus	87.5	24
GR 128	40	ELFTSGTIARSHVRSLSHFQGFRRDHIEMLLSAQTFFVQQLV	pH-sensitive K channel - Mus musculus	62.5	40
GR 129	27	VSARKNEEGEEVAYERFKEAITFPHE	endo-1,4-beta-glucanase	40.0	30
GR 130	24	ETPGYTNHNEKSNCRKVPILTEL	M protein - Streptococcus sp.	40.0	20
GR 131	38	PITIIFODLISHMEMFSDIYKIWEITNGICLEVEQKML	Ig-dependent histamine-releasing fact	77.1	35
GR 132	45	GEKPSNRKLRDRTNLELSEFEIEMINMLKTLIENIDNMEQRMDN	HYPOTHETICAL PROTEIN MJ1417.1	38.7	31
GR 133	15	NHTVSADKNDMKENN	P2Y PURINOCEPTOR 8	90.0	10
GR 134	41	MSAVTAVFHNTGFTSCSKLVKKKINDMKIGKBEIKQSLLP	ATPase 2 [Plasmodium falciparum]	31.8	44
GR 135	30	IVVOIWLRGYSLPTPALYNGKDKVNYLPAP	UROPORPHVRINOMEN DECARBOXYLASE	40.9	22
GR 136	38	MKQEDTQKNYFCGRRVTYKQESWESCFITPTLLDNEE	ORF 381 - Pyrococcus horikoshii	32.4	37
GR 137	20	EIVGGDWICALSNFKATQSMK	lac repressor	53.3	15
GR 138	11	HKKNINVVVEKA	iron	54.5	11
GR 139	40	MISEDVYGFKSTRGGLVIVNIVRSLPAYIFLPRCMGRCDH	(-)		
GR 140	36	SYPLKARMVKIGSLKERKAFNSHVIKDREIDRKEGK	CYANELLE 50S RIBOSOMAL PROTEIN L35.	37.9	29
GR 141	33	EKRKSNPKRSKGYQDDDEERCQDNCVQSQKS	hypothetical protein 1	61.1	18
GR 142	45	YNLTKVWSTDQQHKHYLGACQQCKTSDLRSDDLNLQNLDFNKIPQI	ALDOSE REDUCTASE	35.1	37
GR 143	55	YISIIISQEKFKRERRRRLPKKTKIPDDLNAGKILEDSTKLELGEHVLQMSKP	Mouse endogenous retrovirus in Fv-4	30.5	59
GR 144	69	MTDFSIEDDEKDAIKTKPSYLGCCCKQQVSPHSPTROIHNSQEVGIRAEFE	locus		
		GFCLKVSMEWESQSKDG			
GR 145	19	MQMKANYVQKQEQECVAL	DNA polymerase	73.3	15
GR 146	32	SGDWEIHCQGGTSGENLLAGDLSLQNTVEVGIP	PXNC [Leishmania major]	52.6	19
GR 147	18	MEPGGTGIHIRYKAVKAV	RING finger protein [Rattus norvegicus]	64.2	14
GR 148	34	FPYVYTIYSSKFPFLGSSHEDDDVQKRQVTFLLRPH	40S ribosomal protein	42.8	28
GR 149	36	MSALFCKLLKDKDQSLFVSPVILSPVLCTQQMTN	hypothetical protein ybcP	50.0	24
GR 150	51	KQPQAGSPGGIPEKGTVAEDDSSMPVTAPEDELTVGQDVEVEDSDIDDPDS	ethylene-responsive LEA-like protein	38.6	44
GR 151	38	LNYSQACECIKVKMTDQIPNNLVSFVVKCRVGVGNKK	major allergen:ISOTYPE=Par j 1.0101	48.3	31
GR 152	36	MKTADNDEIYYWCSTHLLPTLQGPLQAPAPNKAVQD	C-TERMINAL BINDING PROTEIN 2	52.3	21
GR 153	33	MLGISAMNIIICIGESEVSAGLEYWEDTGRAEFK	replication initiator	39.1	23
GR 154	11	DFTNNIEQSGL	unknown [Saccharomyces pastorianus]	80.0	10
GR 155	43	CFKVMNDYKGTIEFFYEHELIAAFYLFQSFQKQCYVEYRPC	ORF 10 - Borrelia burgdorferi	27.9	43
GR 156	14	INEVKELKYGKIVV	PAR-E [Mycoplasma genitalium]	63.6	11
GR 157	36	KIERGCEGSNAGEFFILNIVVVKEDDHMEKTPLEEKAO	hypothetical protein 31	39.2	28
GR 158	50	RCSLHQIATAVGQGEWHWRFKAAVFFCFEFTSFRDKLKLKPGTHLIFGSYE	C08H9.11 [Caenorhabditis elegans]	34.0	47

Table 1b

exon name	EST	Exon trap	size (aa)	Predicted Amino Acid Sequence	Gene Name	identity (%)	overlap a.a.
GR 201			42	FFFKMEDPRSCLNADGKKIKFKGRIIDVIRTQVGVGDKN	(-)		
GR 202	+		43	SNPSKTFLEFRVQEVYTKKFRNPNFFLSDKDDIYPVDVTSID	Streptococcus thermophilus bacteriophage	41.6	36
GR 203			51	SCLPFCPQQDSNLIEDHKTFIPGSIPLPHTLEBGMHLREVKKLNLRQACWV	HIV tat protein	37.7	45
GR 204			39	RETGQTSYKGTAVDLIIDEIATDRSQRTVGYIFKSERK	Homo sapiens protein	38.2	34
GR 205			47	GKEEYMLNHAYAILPLEKKTFTFYIFQSNEMMYKVLKVKNPIM	hypothetical protein X	28.1	32
GR 206	+		36	ELNQMLAETPPVWKGSSKLFRLNKEGELTILVLF	Venezuela tomato geminivirus AC1	47.2	36
GR 207			69	VGHYLCLEFRNISQVAVANVGGPPFVFLPPLVPSKEKIQCFEKGTSTE LLFEHGFQVFFCPSFLL	Pelvicachromis pulcher unknown protein	29.7	47
GR 208			40	CPTVVAALMVGHVMIVDQMPFLTMEISWVRQGLLAST	(-)		
GR 209	+		86	SSLFQEKLTSECFREQFEENWYNTYSSNIYKHGDTGRYFVALNKDGTPRD GARSKRHQKETHFLPRVDPERVPPELYKDLLMYT	Xenopus laevis fibroblast growth factor-9	88.4	78
GR 210		+	18	GEINPKAIFLICKMKVFG	Sus scrofa ovarian steroid	44.4	18
GR 211			32	HSSALQNDGEMAVFNLTIDIDVFLPSENGQG	(-)		
GR 212	+		60	AIGRTDIEDLDIYATSRERRRFLFASICEGQLFMPYDFILAVTTDEPKGK	Caenorhabditis elegans C56A3.6 protein	45.0	40
GR 213	+		45	IPSLFLAEPHAGFRJAFNMFDTGNEMVDKKEFLVVCILDAALY	Caenorhabditis elegans C56A3.6 protein	52.3	42
GR 214			73	FPSLLHFCHEMDNLQTEVLEIEFLSYNGMNTISEEDFAHILLRYTNVENT SVFLENVRYISIPPEEKVSNPHI	Caenorhabditis elegans C56A3.6 protein	38.8	54
GR 215			83	ILFLCNIIYLDEEKRAVYVATGLKFSPLVNTVFKIFVDKDDQLSYKEFIGI MKDRLRHGRFVNLHNLNLLSFIKSEGNQSI	Caenorhabditis elegans C56A3.6 protein	50.0	52
GR 216			37	ITDFEFSFFQLNLEDFAIALNMYNFASRSIGQGG	Homo sapiens atopy related protein	40.0	35
GR 217			47	QLYYGSPKQNDVHKEIYCKKSAHVIMEADKSQDLQGEVASWRLR	(-)		
GR 218			63	RSWAGCGWKMAPSGPSSARRRCRRVLIWIPVVFITLLLGWSYAYAIQLC IGCAPRRGAP	(-)		
GR 219			28	LDLFTSIATPEEQSGAEFISKIYTKQ	Feline herpesvirus 1 glycop	50.0	24
GR 220			52	THLDDELCSRYCARRWLHDVQKNPNQTKNELAFVKFIVQIKKMSAPTALI	Caenorhabditis elegans F10F2.5	57.8	19
GR 221			26	KHSMULKHRKLIKIEKYINEVGEVSKD	Rattus norvegicus deoxycytidine kinase	44.4	27
GR 222	+		52	FLYHLELEAFRGPVFRHGTDLNGFSLGFSNNMRHVFGDAKQYLLPIFSRYI	(-)		
GR 223	+	+	35	GKFFKIFSCSAEIQKDVEYRPLPFTINNLTIINM	Beet yellows virus protein	27.2	33
GR 224			58	YIKRGLIGSGFCWLYKKHDIVICLASELELLPVEGKAGAGTLHDESKSEEM	(-)		
GR 225	+		56	HFFFCVSSVQLTDMNEQEEVLEQLFLPQLKQIITDKDDLKVSIEELASM	(-)		
GR 226	+		32	YELLTQMKSTFEKMKQRQHELSEVRLFIFFPL	Azospirillum brasilense unknown	44.0	25
GR 227	+		39	TEKFFPVCSSNYCKMYRQKLIYHRNKIVKYLEVPMK	S. cerevisiae chromosome IV ORF	35.2	34
GR 228	+		67	IKVVPDFLQSCSASALQARLKVAAHEAEESDNIADDFLEGRMEIDDFLSS FMEKRVICNTRQLRT	Caenorhabditis elegans CD4.4	25.8	62
GR 229			49	LARTRHVIRNMLGPCEQQDLCRLVSRKVKKNDIQAKAVRGLLEPRS	Ost oncogene - rat	35.4	48
GR 230	+		41	FFEDHDDAKYCGHLYGLGSSVYQNGTGNAYEEANKQS	Mus musculus mCAF1 protein mRNA	97.5	41
GR 231	+		73	FYSYDFGILIKILITNSLPERELDFEILRFPVIYDVLYMKSKCKNLKV RLRMALSTAHCSDQVIMRA	Mus musculus mCAF1 protein mRNA	98.0	51
GR 232	+		64	ECCLSLSASREDMYAQDSIELLTTSGIQFKKHEEGIEIQTQYFAELLMTSGVV LCEGVKWLSEFHR	Mus musculus mCAF1 protein mRNA	100	53
GR 233			38	GTMNLRGAMAPAGVRLAEEGAWRRIGGWVTGAGLSR	Homo sapiens PFKM protein	58.8	17
GR 234	-		35	TTEFLMGTWLYIFDEIYPKKDKCKIKKKNCCIQ	hypothetical protein PH0080	52.6	19
GR 235			46	FSSFSYLVFISKQLIISKSVTIRKFTAPPCKNGYEEEVGRQSH	Mitochondrion Strongylocent	35.0	40
GR 236	+		49	NTEFFPHSYHKMSSKMLFENFLIYVSLKVAYSKNLTLSEAFYLEM	Caenorhabditis elegans W03B1.2	31.9	47
GR 237			43	NHYHNIIILDKPKNKKLNGYKSPKIDTSRLRVSELMKETEHE	Pasteurella multocida UDP	51.8	27

GR 238	+	128	KGSLMSLFLADGSCDPVVLHRLPTRLIQNGSENAYNNDISILNLEIKSLGN LDQHDFCDLIRKGSISHSNDALITASEAGPEAALISSVLVPKDCDKHKRAE GTKPTSLEEQAAQPOWLFIFQIP	Caenorhabditis elegans K08F4.9	28.2	85
GR 239		119	LEKIQVQLNCTKVKQKSEPSKSHSFGSTSDNSIANTPDYSGNMKSFSPRS PSQGEDSALILTQDNLKSSDPDISANSQESGVEDLSCRSPSGEHAPSED	Myotubularin related protein 7 (mTMR7)	66.6	54
GR 240	+	70	CEFSDCVKVSRRAILSRFWGMYNRFKGMQRPQSVTDYLMVAKETQOL EELEALEEVRHTCFVNL	Myotubularin related protein 7 (mTMR7)	88.3	43
GR 241		76	KVSNALFYRGNLDGDPKIEISPIDQFTECVQWLMQFQCAFENRFLIHI QHIIYSQFNGFELNSQKERRELK	Myotubularin related protein 7 (mTMR7)	94.2	69
GR 242	+	17	VLEIKDWISFGHKFNHR	Myotubularin related protein 7 (mTMR7)	100	17
GR 243	+	51	AVSEEGASVIVHCSGDWRDTAQCVSASILLDPHYRTLLKGFMSVATGRFP	Myotubularin related protein 7 (mTMR7)	93.3	45
GR 244	+	53	SICRSQPLSGFSAKLEDEQMLQAIRKANPQSDFYVVDTRPKVSVTVAMH	Myotubularin related protein 7 (mTMR7)	95.6	46
GR 245	+	53	QCDSYPTELYVPKSAHAHIIVGSKFSRRRRFPVLSYYKDNHYSLEACFC	Myotubularin related protein 7 (mTMR7)	94.7	38
GR 246		57	TQCSGEGRNPHNWCVHRGEASTDASLAIGDWAQVVEFLVAHAKPSKSTY	Myotubularin related protein 7 (mTMR7)	39.5	48
GR 247	+	64	NFILPFLAVKYEELYCFSPNPLDKERQGVWLDLSEETRMGLPNHYWQ LSDVNRDVRVSA	Myotubularin related protein 7 (mTMR7)	46.8	32
GR 248	+	62	LLHSQISTIEKQATATATGCPVLIRCKNFQLIQLIIPQERDCHDVVISLIHLA RPGRAEELGI	Caenorhabditis elegans F53A2.8	46.9	49
GR 249	+	49	IPNSMERYQQTSTSIKTIQENMTLPNELRHQETILEKQRYVTLQTDNS	(-)		
GR 250		53	DISFHSLGPDYDSIAEMPPVLYLHVVISPARAHWILQELSYHHNSYFCY	(-)		
GR 251		40	FFGHLENPDVTKEIFTLRQNELECTVYTKSTCLQIFR	(-)		
GR 252		73	DIYLQDGIKLLLTVYSREVTGGSINKRAYPILHPLSQSAEMQCIQDMEQK LISLQAFQLVDIYSVFFHLGV	(-)		
GR 253		67	AACTPHLSRRRPHKQGPLCVSENSGRIVLRRLPLRESEQSPSPRQPAEAR RGGAGGRRRGGGLS	Caenorhabditis elegans C18D11.4	34.0	47
GR 254		55	LADKLTAKIIANKLVFNIRRYDDMKKFEFFISSITLHSMASGKRHKRCQSSN	Caenorhabditis elegans K07D4.5	39.4	38
GR 255		60	HIHCAFYTSVLMEEETGSLCNKVKQKQVSEYSLPNNDDLEYTRIRIHS	(-)		
GR 256		53	KYTYTHKLELNMKFLTDSRRNTASYQRYAIOKATFNINETEIVCIKELNG	Arabidopsis thaliana R2R3-M	36.6	30
GR 257		48	KRRFYFHSKAYHRISAFIFITVVLRFVVLHKEVQESAEHRP	Schizosaccharomyces pombe hypothetical	23.9	46
GR 258		59	YFTGLFGRNCLCFWDAVNTLILRALISFLSIVFMGNSLWKNVNFV	Homo sapiens gastric mucin	46.3	41
GR 259	+	136	QLLTSLLLRSPSDVRMIPCAAITFARCLIRRIKIVTLDSDTKLCRCLS TMDLIALGVGSTITLGGVYVIAGEVAKADSGPSIVVSLIAALASVMAGLCYA EFGARVPTGSAIYLYTVVGLWAFITGNL	cationic amino acid transporter 2 (CAR2)	100	119
GR 260		63	LKGFVSVGLLSFGVKESAWNKVFTAVNIIVLVLFVWVAGFYKGNVANWKISE EFLKNISASAR	cationic amino acid transporter 2 (CAR2)	100	56
GR 261	+	44	LHRSFLLPLALMAFLFDLKKALVDMMSIGTLMAYSLVAACVLIILR	cationic amino acid transporter 2 (CAR2)	100	35
GR 262	+	74	LIFFPAPTGLIYFSYGRHSLEGLHROENNEEDAPYDNNVHAAABEKSATIQ ANDHPRNLSSPFIHEKTESEF	cationic amino acid transporter 2 (CAR2)	100	65
GR 263		42	PPLRPRPAQPPRSCAPRSEVPEMKVWLLGLLLVHEALEDEG	(-)		
GR 264		45	MHFFGTAGTMTAVRFLNVLVSWVHPVVYTRDRHWVCVLLWCVL	chemokine receptor CXCR3	35.2	51
GR 265		46	PSRRIFLPTRNQNLMPGMAKSGKHKFKWFLDPCLYNTDVTVG	synaptic glycoprotein SC2	31.7	41
GR 266		96	HFFSGRGLDFDGLAHFLIYCVIVVSKIEDNSFCIVVIFISKEIDNHLQRH AVNPVPEFQKVPEDVYGLRQLVLELCRIQLRQRYKQPHRGSKT	ribosomal protein S3 - maize chlorop	31.8	69
GR 267		37	DIEIYESKNLEERKMARLEITEVKYVLEAREQHPWH	(-)		
GR 268	+	50	IHTMEYAAIKRNKIMSEFATMMLKLEAIIILSKMQEOKTKDHIFSLISGS	(-)		
GR 269		50	ISSASFLVLENDITPSTIKSALGATPICSEKVPYLPYLPYIILSCKQG	Arabidopsis thaliana T7I23.	36.9	46
GR 270		46	HALHVSIPGSESEELYAKQGLFINSHDVCVIRRRQRFKYPKPKITADM	aminoglycoside 3',9'-adenylyltransferase	43.5	39
GR 271	+	22	RGLFGSSFCRLYKXKHTNICFQ	Bacillus sp. cytochrome c _o	50.0	18
GR 272		53	LSSVLTTHLPPIIPKYVAVSQLSRAGACFSDDVTHREVLMSFCYMVRLLELPD	probable serine-specific protein kinase	35.1	37
GR 273		42	VLGVFDKSNFTLYYSMTSEKDWGKSAQRTVEPSPVATGGPQPG	(-)		

GR 274	+	109	RISLSLREAVTGOHLPKNKRKPEGENRIKPTNKVKYKPKIPKMKDRDSANSA PKTQSIMQVLDKGRGFQKPAATLSLLAGQTVLRCCKGSRIGWSYPAYLDTFK	PDGF-receptor beta-like tumor suppressor (PRLTS)	100	99
GR 275	+	67	HVFSIPVQKNERYGQLTLVNSTADTGEFCWVQLCSGYICRDKDEAKTGST YIFFTGKILGALMEL	PDGF-receptor beta-like tumor suppressor (PRLTS)	98.1	53
GR 276	+	116	ACVFLPFAEKGEFLVPSYFDVYVLPDRQAVVPCRVTVLSAKVTLHREFF AKEIPANGTDIVYDMKRGFVYLQPHSEHQGVVYCRAEAGRSQISVKYQLLY	PDGF-receptor beta-like tumor suppressor (PRLTS)	92.6	109
GR 277	+	65	TLPVSCLLPVPSGPPSTTILASSNKVKSGDDISVLCITVIGBPDVEFTWIF PGQKVASVPASQP	PDGF-receptor beta-like tumor suppressor (PRLTS)	100	47
GR 278		49	RKHILFFHRKQKRIPTIMDFIPEGNKKTRDSKINVEYQKITGPKGFH	Seed imbibition protein	45.0	20
GR 279	+	79	VHYISFSKGLNSGSGYGLKLSANVISINRLEVFHCNDTILGDSRGV SQSSVDQPPSVLDRHRPFLIFNDTHKD	Caenorhabditis elegans R1IA5.4	29.5	71
GR 280	+	72	HGFWFFSRQLSTEQAVLQESLEKESKVNKRKLSMENEELLNKLHNGDLCSPK RSFTSSAIPLOSPRNSGSP	Homo sapiens KIAA0774	57.1	63
GR 281	+	43	INVGILVYAKVDNNTALVDKLRFOQENEELKARMDKHMATSR	Homo sapiens KIAA0774	40.6	32
GR 282	+	62	TMWKNSSILLCPFQKNPQIMYLEQBLESLKAVLEIKNEKHLHQDIDKLMKMEK LVCFLQNERG	Homo sapiens KIAA0774	43.4	46
GR 283	+	86	FPLLSPLFSVSTTCEKLEKARNELQVYEAFFVQHQAEKTERENRLKEFY TREYEKLRDTYIEBAEKYKMQLOEQVCAFAARAC	Homo sapiens KIAA0774	33.3	75
GR 284	+	41	KQINDLKSENDALNEKLIKSEEQKRAREKANLVSCVCSFIS	Streptococcus pyogenes M protein	42.4	33
GR 285		44	ACLPGVFKHTINGLLQAVTVIOFTCKVTVVRHWGECFTCALIS	(-)		
GR 286	+	90	FVEMLIRFAVEKSRQKNPSLQIQPTAPALPPEKTLLELTQYKTKCENQSG FILQLKQLLACGNTKFEALTVVIHLLSEVRRCILERA	(-)		
GR 287		46	KPLESSFLQPIYISVQCYHLTDEYPELTKTDFDENDIVAYH	Homo sapiens 280 kda protein	32.5	43
GR 288		43	QKVPQREVLMTTRTPQKRSFLGDVSTCLCKYHKMSVASGGD	(-)		
GR 289		57	INYNYYHIEFIIITILMEEELIKIQFKLHFTNEKPLRNQLHDVAKSWNSR	Anguilla japonica ventricular	31.5	38
GR 290		29	AAFQVCCRIKDATDDMEMNKRNCIPVK	Mycoplasma hominis orf499 protein	47.6	21
GR 291	+	14	VDDDKKEFGANFDI	hypothetical protein RP413 - Rickett	53.8	13
GR 292		71	SONHKHDEPRFFMSYYHDLRLTLTWYEEPESEEKLLCGQRTTKCVCFGKTT LMLKNDAEKPVTLGGSLN	(-)		
GR 293		33	HYPAQFWATMAIIFNVNGPVKFLSTECGVNGDS	hypothetical protein yxIE - Bacillus	30.4	23
GR 294		41	GCTHQQKRTDENILKESYKQTOGCGRECWIPODMSPFKR	Arabidopsis thaliana T06D2	38.4	26
GR 295		49	LPSQSSNFSEVDGRLISKPSHDQOSPALISKITOPTCRFVRSNKWLLI	(-)		
GR 296		55	INTHWSSRLICTDLENVANGDLSIVQMRKLGFRMIEVVVTINNLKKSRLR	14 aminoglycoside 6'-N-acetyltransferase	41.0	39
GR 297		28	FSFELFFHTREMGGDVMRCCKDQNLVN	Oryctolagus cuniculus ryanodine receptor	43.4	23
GR 298		54	EKIHFATVCRDKEKVVLLIIVSVPKPOLGMWNEDSGAHRRAIRNYVLHM	(-)		
GR 299	+	62	VFFFLPSTRCHSANLNGVYSGPYTAKTDNGIIVYTWHGWMYSLKSVAMKI RPNDFFIPNVI	(-)		
GR 300	+	67	NEYELNIGESYGTAGDSLGNFHPVEVQWASHQRMKFSTWDRDHDNVEGNCA EEDQSGLWENRFDVIL	fibrinogen-related protein (HFREP-1)	98.4	63
GR 301	+	43	LLVSFLEDYTLKIDLADFEKNSRYAQKFKVGVDEKVLQFSKK	fibrinogen-related protein (HFREP-1)	100	30
GR 302	+	59	GWVVVALVQMEFALIGFHTFFRGMKDYENGFNGFKHGSEYWLGNKLNLFH	fibrinogen-related protein (HFREP-1)	97.1	35
GR 303		64	WLFVFFVFNQDYLSLNNLQNVHVKDRKMFNGKTKHRQCSPPVDNKEQKE NKTLLVFWNLSE	BARBA DNA GYRASE SUBUNIT B	45.4	33
GR 304	+	49	SSVNLFCRELIVKSSMAGGGPFEDGMNDQDLPNWNSNENVDRLNNMV	Centrosome autoantigen (PCM-1)	100	44
GR 305	+	99	RSIGSDSGRATAANNKROLSENKRFNPLFMQINTKSKDASTSPENRETI GSAQCKELFASALNDLLQNCQVSEEDGRGEPAMESQVITFVFAH	Centrosome autoantigen (PCM-1)	96.7	93
GR 306		73	FLKKIFPLQIVSRIVQIRDYITKASSMREDLVEKNERSANVERLTHLIDHLK EQEKSYMFKLKKILVLSIRLLK	Centrosome autoantigen (PCM-1)	100	48
GR 307		57	MIYTAKIFQARENEBEDVRTIDSAVSGSGVAESTSLNIDVQSEASDITVSGF	Rattus norvegicus zinc finger protein	43.1	51

GR 308	+	79	DTSIVGLIFQARDPQQEPEEIEJENLKKQHDLLKRMQLQQEQQLRALQGRQAAL LALQHKAEQAIAMDDSDGMSQSENILF	Centrosome autoantigen (PCM-1)	84.0	27
GR 309	+	47	SVVTFPPVVAETAGSLSGVSTISELNEELNDLIQRFHQLRDSQVT	Centrosome autoantigen (PCM-1)	100	26
GR 310	+	82	FLFFFNKPPAVPDRNRQAESLSLTREVSQRKPSASERLPDEKVELFSKMR VLQEKQKQMDKLLGELHTLRDQHLNNSSE	Centrosome autoantigen (PCM-1)	97.2	31
GR 311		68	IVYINFLFIRKLNVEKRRLNELRELVHYEQTSDMMTDAVNENRKRDEETEES EYDSEHENSEPVNIR	Centrosome autoantigen (PCM-1)	96.6	39
GR 312	+	93	YMSADCRYNREGEQEIHVAQGEDDEBEDEEEAEIEEGVSGASLSSHRSLVDE HPEDAEEFEKINRLMAAKQKRLQLDLVAMVQVNIAWSFKN	Centrosome autoantigen (PCM-1)	98.7	26
GR 313	+	69	DCENVMLFQDDDAAGVISASASNLDDFPAAEDTKQNSNTRGNANKTQK DTGVNEKARYVKLLAFI	Centrosome autoantigen (PCM-1)	89.6	36
GR 314		57	QFNCREKFEYEQRLQQOQRELKQEQERKKLIDIQEKIQALQACPDQLQVIMK	Centrosome autoantigen (PCM-1)	84.9	27
GR 315		82	TSVLLITRIMATWGGSTQCALDEEGDEDDGYSLEGIVTRDEEEEEEQDASS NDNFVSCPSSHVSNHNSYNGKTKNRLVSVF	Centrosome autoantigen (PCM-1)	95.5	45
GR 316	+	44	IVSRHISEHKEGNVKSVMWIASNSELTPESELATITDDVS	Centrosome autoantigen (PCM-1)	100	29
GR 317		46	ETFEKNFERETHKISEQNDADNASVLSVSNFEPFATDDIGKQNCI	Centrosome autoantigen (PCM-1)	100	19
GR 318	+	64	SKNIFVPVNGNTVIHLQALARMREYERMKTEAESNSMRCICRIIEDGDGA GAGTTVNNLEGI	Centrosome autoantigen (PCM-1)	94.5	55
GR 319	+	58	DFSQEHMDEVCSQLITSVRRMVLTITQONDESKEFVFKFKQLGSLQVRV	Centrosome autoantigen (PCM-1)	95.8	45
GR 320	+	74	DSLAKETAGRKLKDCGEDLLVEISEVFNELAFFKLMQDLNNSITVKQRCKR KIEATGVIQISCAKEVNNVHFDV	Centrosome autoantigen (PCM-1)	91.7	73
GR 321		58	DKDETEVKTQITSEVYDQPKNVRSDISDQEEDEESEGEPVSIKFKGVSLS	Centrosome autoantigen (PCM-1)	59.2	63
GR 322	+	28	LQRDFKTAESKNVFLEREAATSKSKSK	Centrosome autoantigen (PCM-1)	53.8	26
GR 323	+	46	YVFFFITLQEAESGNIQSQBDEDFVKEVDLPLKLTIIYSEVFSCL	Centrosome autoantigen (PCM-1)	100	31
GR 324		52	CLCCLFLVRCRQLLKENQNHKQMIEMAVLNKNGGNGVQNGRGLDVGRRG	MAN RETROVIRUS-RELATED ENV POLYPROTEIN	47.7	44
GR 325		68	APPNKDIPAGQSSTFKSRANTAVNESLFTKQGGSSVMHCNCDSSVHHKVSF NSNTKDNITDNNVKKLPV	HIV rev protein	32.7	61
GR 326		42	RLCIFLNCFPAGQGEYDHPENMTEANNNNYNNVNCIEBTK	LAFK CALCIUM-TRANSPORTING ATPASE	38.4	39
GR 327		33	PHLNIDDSYGLNVSQNSGVSVLVRGAFFKR	Saccharomyces cerevisiae Ypr144cp	46.1	26
GR 328		36	RYNLSIQAFQTEDDKANGSATEFENPVYLITCSGK	HELPSY TRYTOPHAN SYNTHASE BETA CHAIN	36.1	36
GR 329		30	VVKSIRKLSLEIDGPKAENPEVLLHCE	Saccharomyces cerevisiae pyruvate kinase	44.4	27
GR 330	+	49	MELYSYEEAKNLTTKTILAPAYFILGNGQSGEGCVITRDRKESLDVYE	Homo sapiens PHP protein.	100	45
GR 331	+	45	AFINIFVQPGLLGNFPGPEEMKGIAAVTDIDIGKVHLEALKK	Homo sapiens PHP protein.	93.1	29
GR 332	+	35	FSKHFIYFAKGYSEFSEYDKYIRAKWKSYAGGG	Homo sapiens cuniculus angiotensin	52.6	19
GR 333	+	55	TEVCFSDLFIMHLAEILDYENIFRFKCPSLDYQNHKKEKDFQPDVAHAHGN	Oryctolagus cuniculus angiotensin	41.8	19
GR 334		61	KESVHLSQAVTSLGAVLTSRAESGINKERKQMIKQCHKHVDSEVHKCC GRKGGYEYF	Homo sapiens KIAA0273 protein (-)	41.8	55
GR 335	+	69	LRDVSYHCFVGLNLFQFNGLFVVSAGGSIPSSGTKTSRPAEPMNGARCKNF ASGSPGVMSVSGATSIST	murine herpesvirus 68 unknown protein	42.5	40
GR 336		48	CYFLITGCNQHNVINVMQDQIILWKQNVIKFSSNKITTSQRKVDIPLR	Tetrahymena thermophila P-type ATPase	31.1	45
GR 337		100	KKPTMDHRTSPLHLQPKRFRKILSFTKAAVQKDISSETDDDFNFIHMHNA YTQSMLYKCTLMKSYLQEMHKINQLESYTIYTSHTGSDVTVKYKAR	(-)		
GR 338		63	INKVRDRKGDITIDTTEIQKIIREYEQLYDNTLESGLGEMDKFLHTCNLPLL NHEAIQNMNRQ	Plasmodium falciparum malaria antigen	57.1	63
GR 339		23	AKKICFSSNRIREINEVKHEREM	Vesicular stomatitis virus orf2	61.5	13
GR 340		44	GLPVGLGPVSYVMLKSLKHKTLDSNAEKIKITYPKAFNMIQSS	Caenorhabditis elegans C44C8.1	51.6	31
GR 341		39	GLLSITPHRNIKWNNYPCKTPSKELKKSGLSQYLVA	SCHPO HYPOTHETICAL 40.9 KD PROTEIN C31A	41.4	41
GR 342		39	LVIWNPDETQKTDTERERTKNNIGLKRGIISQIQKQ	Coxiella burnetii protein	53.8	26
GR 343	+	72	LPSSSFASSRLPVLLFDLFDLQALTFHITHTGTSVRLPESRNQVKYKSTVLP GGMGIIGQYNSPPFTVRYE	(-)		
GR 344		71	NSCHQTEMTGILLEDLQREGTESEQREHTEGGLRGGLTGLHGSAAYLVL LCPSSSRGTGELNWNYGATH	(-)		