nonmotor tasks. Eyeblink conditioning requires that an animal not only learn the association between a neutral and an aversive stimulus; it is only adaptive if the animal also learns when to expect the aversive stimulus. More generally, the consolidation of motor skills centers on learning to specify the precise timing between successive movements. Cerebellar ataxia is characterized as a disruption in the timing of these events (*3, 26, 28*), rather than as a loss of the conceptual knowledge for actions observed in apraxia. Similarly, comparing the duration of two successive events requires a judgment of whether the second event occurred earlier or later than expected. Our results extend previous theories concerning the role of the cerebellum in temporal processing, indicating that this function is limited to tasks that require an explicit specification of the timing of behaviorally meaningful events.

References and Notes

1. R. B. Ivry, S. W. Keele, H. C. Diener, *Exp. Brain Res.* **73**, 167 (1988).
2. R. B. Ivry, H. S. Gopal, in *Attention and Performance XIV*, D. E. Meyer, S. Kornblum, Eds. (MIT Press, Cambridge, MA, 1993), pp. 771–802.
3. D. Timmann, S. Watts, J. Hore, *Exp. Brain Res.* **130**, 441 (2000).
4. R. B. Ivry, S. W. Keele, *J. Cogn. Neurosci.* **1**, 136 (1989).
5. J. A. Mangels, R. B. Ivry, N. Shimizu, *Brain Res. Cogn. Brain Res.* **7**, 15 (1998).
6. H. Ackermann, S. Graeber, I. Hertrich, I. Daum, *Brain Lang.* **60**, 323 (1997).
7. I. Daum, M. M. Shugens, H. Ackerman, W. Lutzenberger, *Behav. Neurosci.* **107**, 748 (1993).
8. H. Topka, J. Valls-Sole, S. G. Massaquoi, M. Hallett, *Brain* **116**, 961 (1993).
9. D. S. Woodruff-Pak, M. Papka, R. B. Ivry, *Neuropsychology* **10**, 443 (1996).
10. M. Jueptner *et al.*, *Neurology* **45**, 1540 (1995).
11. R. Kawashima *et al.*, *J. Neurophysiol.* **83**, 1079 (2000).
12. H. Ackermann, A. Riecker, K. Mathiak, M. Erb, W. Grodd, D. Wilddruber, *Neuroreport* **12**, 4087 (2001).
13. R. B. Ivry, *Int. Rev. Neurobiol.* **41**, 556 (1997).
14. J. Jackson, J. A. Michon, in *Tutorials in Motor Neuroscience*, J. Requin, G. Stelmach, Eds. (Kluwer, Norwell, MA, 1991), NATO Advanced Study Institute series, vol. 62, pp. 169–198.
15. R. B. Ivry, *Ann. NY Acad. Sci.* **682**, 214 (1993).
16. F. Mussa-Ivaldi, in *Biomechanics and Neural Control of Posture and Movement*, J. M. Winters, P. E. Crago, Eds. (Springer-Verlag, New York, 2000), pp. 325–333.
17. We focus on measures of variable error, operationalized as the standard deviation of the cycle durations. This measure has been commonly used in neuropsychological studies [e.g. (*1*)] to assess the contribution of different neural structures to timing based on the assumption that lesions add noise to the system.
18. S. D. Robertson *et al.*, *J. Exp. Psychol. Hum. Percept. Perform.* **25**, 1316 (1999).
19. H. N. Zelaznik, R. M. Spencer, J. Doffin, *J. Mot. Behav.* **32**, 193 (2000).
20. H. N. Zelaznik, R. M. Spencer, R. B. Ivry, *J. Exp. Psychol. Hum. Percept. Perform.* **28**, 575 (2002).
21. By "explicit," we mean that the action representation includes the specification of a temporal goal, as opposed to being a conscious representation as this term implies in the memory literature.
22. Materials and methods are available as supporting material on *Science* Online.
23. P. Killeen, N. Weiss, *Psychol. Rev.* **94**, 455 (1987).
24. Variability due to linear drift in cycle duration was removed by detrending (*34*). A regression line was fit to the time series for each trial, and the residuals
25. were pooled across trials for each participant in each condition. The standard deviation of these residuals was divided by the mean cycle duration to obtain the coefficient of variation scores.
26. One concern with our within-subject comparison is that the difference between hands may be related to other factors, such as handedness. However, the lesions affected the dominant hand for five of the six patients; thus, any effect of handedness would work against our hypothesis.
27. G. Holmes, *Brain* **62**, 1 (1939).
28. A. J. Bastian, T. A. Martin, J. G. Keating, W. T. Thach, *J. Neurophysiol.* **76**, 492 (1996).
29. J. Hore, B. Wild, H. C. Diencer, *J. Neurophysiol.* **65**, 563 (1991).
30. S. W. Kennerley, J. Diedrichsen, E. Hazeltine, A. Semjen, R. B. Ivry, *Nature Neurosci.* **5**, 376 (2002).
31. M. T. Turvey, in *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, R. Shaw, J. Bransford, Eds. (Erlbaum, Hillsdale, NJ, 1977), pp. 211–265.
32. N. Hogan, T. Flash, *Trends Neurosci.* **10**, 170 (1987).
33. R. J. van Beers, P. Baraduc, D. M. Wolpert, *Phil. Trans. R. Soc. Lond.* **357**, 1137 (2002).
34. E. Burdet, R. Osu, D. W. Franklin, T. E. Milner, M. Kawato, *Nature* **414**, 446 (2001).
35. D. Vorberg, A. Wing, in *Handbook of Perception and Action, Vol. 2: Motor Skills*, H. Heuer, S. W. Keele, Eds. (Academic Press, San Diego, CA, 1996), pp. 181–262.
36. Supported by NSF dissertation improvement grant no. 0121930 (R.M.C.S.) and by NIH grant nos. NS30256, NS40813, and NS17778. We thank S. Keele and S. Grafton for comments on this manuscript.

Characterization of Mammalian Selenoproteomes

Gregory V. Kryukov,[1] Sergi Castellano,[2] Sergey V. Novoselov,[1] Alexey V. Lobanov,[1] Omid Zehtab,[1] Roderic Guigó,[2] Vadim N. Gladyshev[1]*

In the genetic code, UGA serves as a stop signal and a selenocysteine codon, but no computational methods for identifying its coding function are available. Consequently, most selenoprotein genes are misannotated. We identified selenoprotein genes in sequenced mammalian genomes by methods that rely on identification of selenocysteine insertion RNA structures, the coding potential of UGA codons, and the presence of cysteine-containing homologs. The human selenoproteome consists of 25 selenoproteins.

In the universal genetic code, 61 codons encode 20 amino acids, and 3 codons are terminators. However, the UGA codon has a dual function in that it signals both the termination of protein synthesis and incorporation of the amino acid selenocysteine (Sec) (*1–3*). Available computational tools lack the ability to correctly assign UGA function. Consequently, there are numerous examples of misinterpretations of UGA codons as both Sec codons (*4*) and terminators (*5, 6*), including annotations of the human genome (*7, 8*), where no selenoproteins have been correctly predicted. With 18 human selenoprotein genes previously discovered (*3*), the estimates of the actual number of such genes vary greatly (*9*). All previously characterized selenoproteins except selenoprotein P (*10*) contain single Sec residues that are located in enzyme-active sites and are essential for their activity. Thus, misidentification of UGA
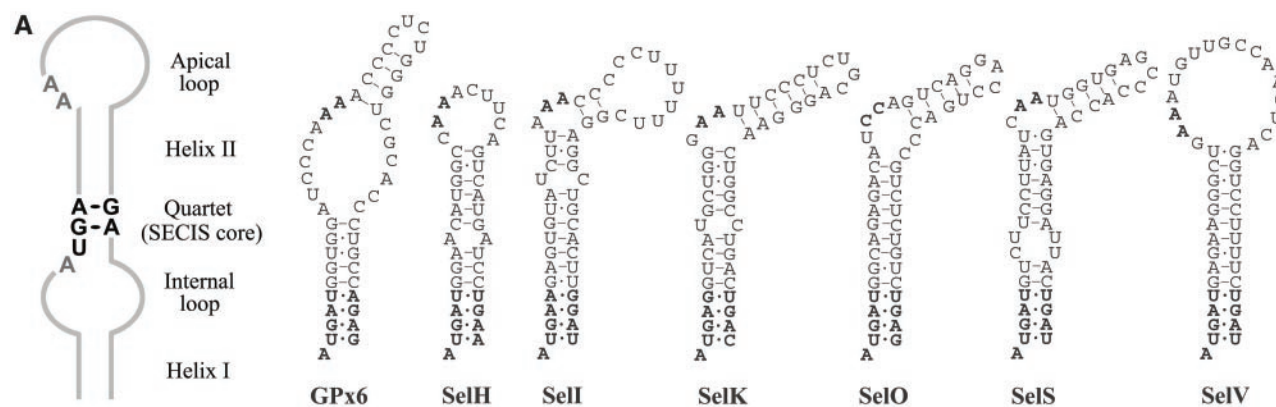
codons leads to a loss of crucial biological and functional information. Sec is cotranslationally incorporated into nascent polypeptides in response to UGA codons when a specific stem-loop structure, designated the Sec insertion sequence (SECIS) element, is present in the 3' untranslated regions (UTRs) in eukaryotes and in archaea, or immediately downstream of UGA in bacteria (*1, 11–13*). Trans-acting factors, including Sec tRNA, Sec-specific elongation factor, selenophosphate synthetase (SPS), Sec synthase, and a SECIS-binding protein, are also required for Sec biosynthesis and insertion (*1, 3, 13–15*). Most known selenoprotein genes have homologs, in which Sec is replaced with cysteine (Cys). However, these proteins are poor catalysts as compared with selenoproteins (*3*).

We hypothesized that the UGA dual-function problem could be solved by identifying selenoprotein genes in sequenced genomes and assigning terminator functions to the remaining in-frame UGAs. The requirement of SECIS elements for Sec insertion and the presence of Cys-containing homologs of selenoproteins suggested two independent bioinformatics methods for selenoprotein identification. In addition, we used an ob-

[1]Department of Biochemistry, University of Nebraska, Lincoln, NE 68588−0664, USA. [2]Grup de Recerca en Informàtica Biomèdica, Institut Municipal d'Investigació Mèdica, Universitat Pompeu Fabra, Centre de Regulació Genòmica, Doctor Aiguader 80, 08003 Barcelona, Catalonia, Spain.

*To whom correspondence should be addressed. E-mail: vgladyshev1@unl.edu

Fig. 1. Mammalian selenoprotein genes. (**A**) Mammalian SECIS element consensus and SECIS elements in newly unidentified human selenoprotein genes. Only the upper portions of SECIS elements are shown. (**B**) Human selenoprotein genes. Proteins are shown in alphabetical order and the newly identified genes are highlighted. On the right, relative lengths of selenoproteins are shown and Sec locations within the proteins are indicated by red vertical lines. The regions in selenoproteins that correspond to downstream α helices are highlighted.

new selenoprotein sequences (16–18), and related approaches have also been developed (19). However, these methods were insufficient in identifying selenoprotein genes in mammalian genomes because of their size and complexity.

Our SECIS-based method, as applied to mammalian genomes (fig. S1), consisted of the following principal steps (20): (i) We identified candidate SECIS elements in the human genome with SECISearch 2.0. This program analyzed structural and thermodynamic features of SECIS elements and was about 10 times more selective (with the same specificity) than the original version of SECISearch (16). (ii) We identified human/ mouse and human/rat SECIS pairs with SECISblastn, a program that analyzed evolutionary conservation of mammalian SECIS elements. This program was based on our observation that human, mouse, and rat SECIS elements in orthologous selenoprotein genes exhibited detectable sequence similarity. SECISblastn provided an increase of about 100-fold in the specificity of genomic searches. (iii) We analyzed genomic sequences upstream of candidate SECIS elements with geneid (18), a gene prediction program that identified open reading frames (ORFs) that had high coding potential and that contained in-frame TGA codons. (iv) We analyzed predicted human selenoprotein genes with mammalian selenoprotein gene signature (MSGS) criteria (21), which screened selenoprotein homologs for the presence and conservation of ORFs, in-frame TGA codons, and SECIS elements.

Primary sequences of more than 95% previously characterized mammalian SECIS elements contain an adenosine that precedes the quartet of non–Watson-Crick base pairs, a TGA_GA motif in the quartet, and two adenosines in the apical loop or bulge (12) (the ATGA_AA_GA pattern) (Fig. 1A). In addition, in mammalian SelM SECIS elements, AA is replaced with CC (22) (the ATGA_CC_GA pattern). The SECISearch 2.0 screen of mammalian genomes using the ATGA_AA_GA
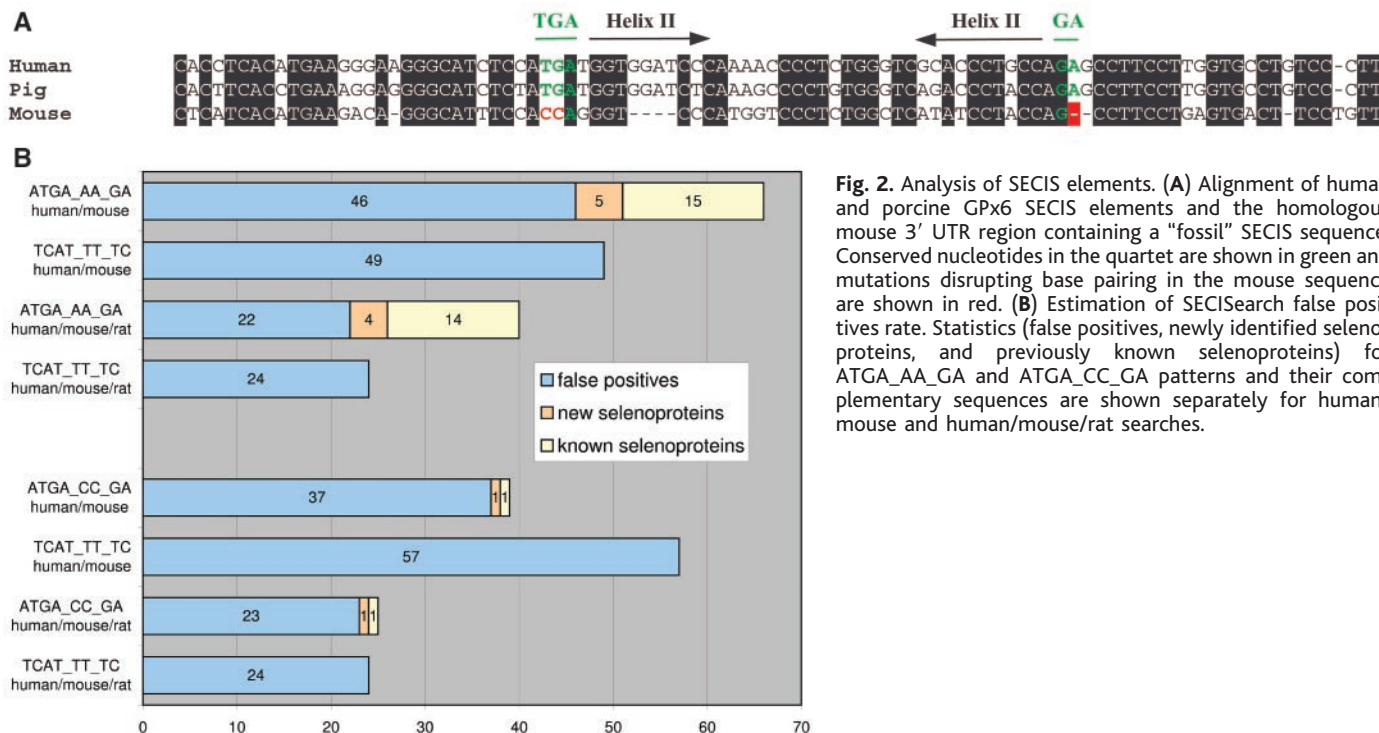
servation that the strong codon bias characteristic of protein-coding regions extends beyond the UGA codon in selenoprotein

genes. We previously developed two computer programs, SECISearch 1.0 and geneid, which were used to identify several

A

B

Fig. 2. Analysis of SECIS elements. (**A**) Alignment of human and porcine GPx6 SECIS elements and the homologous mouse 3′ UTR region containing a "fossil" SECIS sequence. Conserved nucleotides in the quartet are shown in green and mutations disrupting base pairing in the mouse sequence are shown in red. (**B**) Estimation of SECISearch false positives rate. Statistics (false positives, newly identified selenoproteins, and previously known selenoproteins) for ATGA_AA_GA and ATGA_CC_GA patterns and their complementary sequences are shown separately for human/mouse and human/mouse/rat searches.

pattern resulted in 7146 human structures. The SECISblastn analysis reduced the number of structures to 1031 human/mouse and 276 human/rat pairs, and subsequent use of contamination, shotgun redundancy, and repetitive element filters resulted in 56 unique human/mouse and 58 unique human/rat pairs, including 40 structures that were common to all three organisms. The geneid analyses of sequences upstream of candidate SECIS elements and a subsequent analysis with MSGS criteria reduced the set to 20 hits. Among these, 15 were already known human selenoproteins and 5 were novel selenoproteins, designated as SelH, SelI, SelK, SelS, and SelV (Fig. 1B, figs. S2 to S6, and figs. S10 and S11).

A similar computational screen using the ATGA_CC_GA pattern (*23*) detected a single true positive selenoprotein (SelM) and one novel selenoprotein (SelO) (Fig. 1, A and B; fig. S7; and figs. S10 and S11). Only two known human selenoprotein genes were not identified by these procedures: The *SPS2* gene was absent in the human genome assembly, whereas the thioredoxin reductase 2 (TR2) gene contained a SECIS element with a thymidine preceding the quartet, a structure that does not correspond to other known SECIS elements.

The 24 mammalian selenoproteins were subsequently examined for the presence of homologs. This analysis identified a 25th human selenoprotein, designated glutathione peroxidase 6 (GPx6) (figs. S8, S10, and S11), a close homolog of plasma GPx3. GPx6 was not identified in the SECISearch-based computational screen, because its mouse and rat
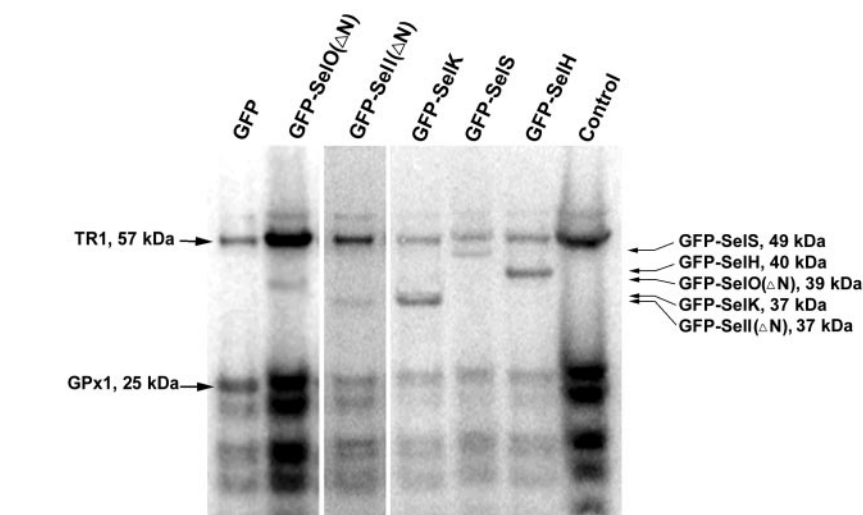
Fig. 3. Incorporation of selenium into newly identified mammalian selenoproteins. GFP-selenoprotein constructs were used for convenient visualization of signals, wherein the fusion proteins differed in size from endogenous selenoproteins. Also for convenient visualization, the N-terminal regions of SelO and SelI were deleted. After transfection into CV-1 cells, transfected and control cells were incubated with [75]Se[selenite] for 24 hours, the extracts were resolved by SDS–polyacrylamide gel electrophoresis, and the labeled selenoproteins were visualized with a PhosphorImager. Locations of transfected selenoproteins are indicated on the right, and locations of major endogenous selenoproteins (TR1 and GPx1) are on the left. The left lane (GFP) shows control transfection with GFP alone. The right lane (control) shows untransfected CV-1 cells. The five middle lanes show experiments with indicated selenoproteins. All five showed ^{75}Se-labeled bands of the size expected if TGA encoded Sec.

orthologs had Cys in place of Sec and the corresponding genes lacked SECIS elements. Rat GPx6 was previously cloned as rat odorant-metabolizing protein (*24*). Homology analyses revealed a "fossil," nonfunctional SECIS element in the 3′ UTR of the mouse GPx6 gene, which contained mutations that disrupt-

ed the quartet and secondary structure (Fig. 2A). We also cloned the gene encoding porcine GPx6 and found that it had a SECIS element and encoded a selenoprotein. These data revealed that Sec, which was initially present in the mammalian GPx family, was replaced by Cys in rodent genes for GPx6.

1441

To estimate the number of false positives in the set of hits selected by SECISearch and SECISblastn, searches were performed using patterns that were complementary to the conserved SECIS sequences. The false positive rate with such patterns should be similar to that in the SECIS patterns, but the true positive rate with the complementary patterns should be zero. The difference between the number of SECIS candidates conforming to the major SECIS pattern, ATGA_AA_GA, and that of the complementary pattern corresponded approximately to the number of identified selenoprotein genes (Fig. 2B). Thus, the ability of our SECIS-based method to recognize known mammalian selenoproteins and to complete analyses of all other candidates indicates that all or almost all selenoproteins common to human and rodent genomes were identified by our procedures.

In addition, neither the SECISearch analyses of human and mouse dbEST and pairwise searches of human/mouse genomes with altered SECIS patterns (23), nor the SECIS-independent searches for Sec/Cys pairs in homologous sequences (see below), revealed additional mammalian selenoproteins. The seven new human selenoproteins were either incorrectly predicted or not detected at all in Celera (8), National Center for Biotechnology Information (7), and Golden Path (25) human genome assemblies and annotations. In new as well as in known selenoproteins, Sec was located either upstream of an α helix or very close to the C terminus (Fig. 1B).

When the SECISearch-based method was applied to other eukaryotic genomes, we found neither selenoprotein genes nor Sec insertion machinery genes in yeast *Saccharomyces cerevisiae* or *Schizosaccharomyces pombe*, or in plant *Arabidopsis thaliana* genomes, whereas we could find only one and three already known selenoproteins in *Caenorhabditis elegans* and *Drosophila melanogaster* genomes, respectively (26) (fig. S12).

GPx6 and SelV were homologs of the previously characterized selenoproteins GPx1 and SelW, respectively, and shared a conserved Sec with these proteins. To validate the remaining five new selenoproteins, we demonstrated the incorporation of selenium into these proteins by metabolic ^{75}Se labeling of CV-1 cells that were transfected with selenoprotein constructs (Fig. 3). Analysis of the expression patterns of these selenoprotein genes revealed that SelH, SelI, SelO, SelS, and SelK mRNAs were present in a variety of tissues and cell types (23). However, the GPx6 mRNA was only detected in embryos and olfactory epithelium (Fig. 4A), and expression of SelV mRNA was restricted to testes (Fig. 4B), where it occurred in seminiferous
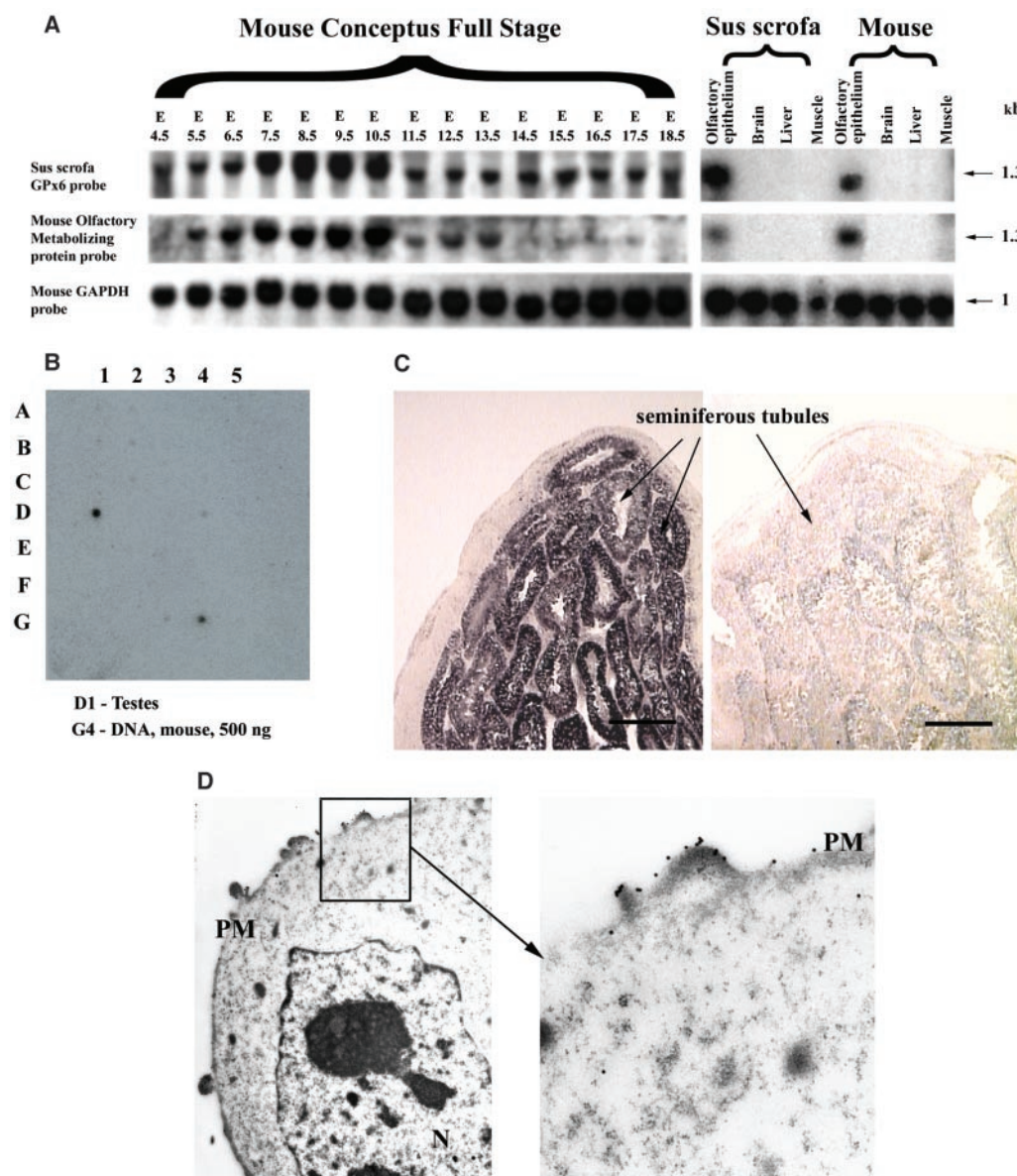
Fig. 4. Expression of mammalian selenoproteins. (**A**) GPx6 mRNA is expressed in embryos and olfactory epithelium. On the left, a mouse full-stage conceptus Northern blot (See-Gene, Del Mar, CA) was probed with pig GPx6, mouse GPx6, and glyceraldehyde-3-phosphate dehydrogenase cDNA probes. On the right, mRNA isolated from indicated mouse and pig tissues was probed as above. We observed no significant cross-hybridization with other GPx mRNAs, which also migrated differently than the 1.3-kb GPx6 mRNA on these northern blots. (**B**) SelV mRNA is expressed in testes. A mouse multiple-tissue blot was developed with a mouse SelV mRNA probe. Northern blots also revealed testes-specific expression (23). (**C**) In situ hybridization of SelV mRNA in seminiferous tubules. On the left, a SelV sense probe was used. On the right, a SelV antisense probe (control) was used. (**D**) SelS and SelK are plasma membrane proteins. A construct encoding SelS-GFP fusion protein was generated and transfected into NIH 3T3 cells, and the expressed protein was detected with antibodies to GFP by means of electron microscopy.

tubules (Fig. 4C). The secondary structure and protein organization predictions suggested that, like all previously characterized mammalian selenoproteins, GPx6, SelH, SelO, and SelV were globular proteins. However, SelK and SelS were predicted membrane proteins. We expressed fusions of SelK (*23*) and SelS (Fig. 4D) containing a C-terminal green fluorescent protein (GFP) tag in CV-1 cells and found that the fusion products did reside on the plasma membrane. Thus, SelK and SelS are the first known plasma membrane selenoproteins.

We next applied the Sec/Cys homology method to the human genome in two different ways. First, we predicted with geneid, and regardless of SECIS elements, all possible human genes that were interrupted by in-frame TGA codons. The predicted ORFs were extended from TGA to the next terminator signal and were analyzed by BLASTP and TBLASTN against all proteins predicted in completely sequenced eukaryotic genomes. This procedure was designed to identify sequences with homology in TGA-flanking regions, which either conserve TGA or replace TGA with TGC or TGT (Cyst codons). Second, we analyzed by TBLASTN all human proteins against all human expressed sequence tags to identify paralogs that contain TGA in place of a Cys codon. These two Sec/Cys homology approaches recognized the majority of selenoprotein genes that

were found through SECIS elements but did not identify additional selenoproteins (*23*), providing additional evidence that all or virtually all mammalian selenoproteins have been identified in our work.

Dietary selenium plays an important role in cancer prevention (*27*), immune function (*28*), aging (*17*), male reproduction (*28*), and other physiological and pathophysiological processes (*29*). Selenoproteins are thought to be responsible for most biomedical effects of dietary selenium and are essential to mammals. Information on a set of human and mouse selenoproteins should provide the basis for future systematic analysis of mammalian selenoprotein functions.

References and Notes

1. A. Bock, *Biofactors* **11**, 77 (2000).
2. S. C. Low, M. J. Berry, *Trends. Biochem. Sci.* **21**, 203 (1996).
3. D. L. Hatfield, V. N. Gladyshev, *Mol. Cell. Biol.* **22**, 3565 (2002).
4. L. Cataldo *et al.*, *Mol. Reprod. Dev.* **45**, 320 (1996).
5. V. N. Gladyshev, K.-T. Jeang, T. C. Stadtman, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6146 (1996).
6. M. J. Guimaraes *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 15086 (1996).
7. E. S. Lander *et al.*, *Nature* **409**, 860 (2001).
8. J. C. Venter *et al.*, *Science* **291**, 1304 (2001).
9. D. Behne *et al.*, *Biol. Trace Elem. Res.* **55**, 99 (1996).
10. R. F. Burk, K. E. Hill, *Bioessays* **21**, 231 (1999).
11. M. J. Berry *et al.*, *Nature* **353**, 273 (1991).
12. R. Walczak, E. Westhof, P. Carbon, A. Krol, *RNA* **2**, 367 (1996).
13. R. M. Tujebajeva *et al.*, *EMBO Rep.* **1**, 158 (2000).
14. D. Fagegaltier *et al.*, *EMBO J.* **19**, 4796 (2000).
15. P. R. Copeland *et al.*, *EMBO J.* **19**, 306 (2000).
16. G. V. Kryukov, V. M. Kryukov, V. N. Gladyshev, *J. Biol. Chem.* **274**, 33888 (1999).
17. M. J. Martin-Romeo *et al.*, *J. Biol. Chem.* **276**, 29798 (2001).
18. S. Castellano *et al.*, *EMBO Rep.* **2**, 697 (2001).
19. A. Lescure, D. Gautheret, P. Carbon, A. Krol, *J. Biol. Chem.* **274**, 38147 (1999).
20. Materials and methods are available as supporting material on *Science* Online.
21. G. V. Kryukov, V. N. Gladyshev, *Methods Enzymol.* **347**, 84 (2002).
22. K. V. Korotkov, S. V. Novoselov, D. L. Hatfield, V. N. Gladyshev, *Mol. Cell. Biol.* **22**, 1402 (2002).
23. G. V. Kryukov *et al.*, data not shown.
24. T. N. Dear, K. Campbell, T. H. Rabbitts, *Biochemistry* **30**, 10376 (1991).
25. J. W. Kent, D. Haussler, *Genome Res.* **11**, 1541 (2001).
26. G. V. Kryukov, V. N. Gladyshev, unpublished data.
27. G. F. Combs Jr., L. C. Clark, B. W. Turnbull, *Biofactors* **14**, 153 (2001).
28. F. Ursini *et al.*, *Science* **285**, 1393 (1999).
29. M. P. Rayman, *Lancet* **356**, 233 (2000).
30. We thank D. L. Hatfield for helpful discussions and Y. Zhou for assistance with microscopy. Supported by NIH GM61603 (to V.N.G) and Ministerio de Ciencia y Tecnologia BIO2000-1358-C02-02 (to R.G.). S.C. is the recipient of a predoctoral fellowship from Generalitat de Catalunya.

Science

Characterization of Mammalian Selenoproteomes

Gregory V. Kryukov, Sergi Castellano, Sergey V. Novoselov, Alexey V. Lobanov, Omid Zehtab, Roderic Guigó and Vadim N. Gladyshev