Compendium of proteins containing segments that exhibit zero-tolerance to amino acid variation in humans

Adam L. Sanders | Jake N. Hermanson | David C. Samuels | Lars Plate | Charles R. Sanders

Abstract

Genetic missense tolerance ratio (MTR) analysis systematically evaluates all possible segments in a given protein-encoding transcript found in the human population. This method scores each segment for the number of observed missense variants versus the number of silent mutations in that same segment. An MTR score of 0 indicates that no missense mutations are observed within a given segment. This is indicative of evolutionary purifying selection, which excludes mutations in that segment from the general human population. Here, we conducted MTR analysis on each of the roughly 20,000 protein-encoding human genes. It was seen that there are 257 genes with at least one 31-residue encoding segment with $MTR = 0$ (1.3% of all human genes). The proteins encoded by these 257 genes were tabulated along with information regarding the sequence location of each intolerant segment, the likely function of the protein, and so forth. The most functionally-enriched family among these proteins is a collection of several dozen proteins that are directly involved in RNA splicing. Some of the other proteins with zero-tolerance segments have thus far escaped significant characterization. Indeed, while a number of these proteins have previously been genetically linked to human disorders, many have not. We hypothesize that this compendium of human proteins with zero-tolerance segments can be used to complement disease mutation data as a pointer to genes and proteins that are associated with interesting and underexplored human biology.

Keywords: database, gene, genetic, genome, intolerance, intolerant, missense tolerance ratio, protein, proteome

Abbreviation: MTR, missense tolerance ratio.

Adam L. Sanders and Jake N. Hermanson contributed equally to this study.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. Protein Science published by Wiley Periodicals LLC on behalf of The Protein Society.
1 | INTRODUCTION

Genetic intolerance analysis has emerged as a powerful tool for studying protein evolution, structure-function relationships, and linkage of proteins to disease. Here we examine an extreme form of protein sequential intolerance by identifying human proteins that contain segments in which genetic variation is completely disallowed by evolution.

Petrovsky and coworkers introduced an approach to measure the “missense tolerance ratios” (MTR) for segments of human protein-encoding genes. For a given gene, this method is based on analyzing the >10^5 sequences for that gene in the gnomAD database and comparing the number of missense mutations present in each 31 amino acids segment versus the number of observed silent missense mutations in that same segment. Coding genes with segments that exhibit fewer amino acid-encoding mutations than expected based on the observed number of silent mutations for that segment are deemed to be genetically intolerant. Intolerance indicates that amino acid-encoding missense mutations within that gene segment are evolutionarily excluded from the human gene pool by “purifying selection.” Because of the limited number of currently available sequences in gnomAD, statistically meaningful single codon MTR ratios cannot yet be determined. However, analysis of 93-base segments encoding 31 amino acids usually yields robust statistics. A segmental MTR score of 1.0 indicates that the sequence of the analyzed gene segment is under no purifying selective pressure, whereas an MTR score of 0 means that the introduction of even a single amino acid-varying missense mutation into a segment is not seen in gnomAD, indicative of stringent purifying selection for variations associated with that segment. Mutations occurring in an intolerant segment of a protein can result in reduced evolutionary fitness through any one of a variety of potential mechanisms, such as triggering the loss of that protein’s native function or inducing the formation of toxic aggregates, as we have reviewed elsewhere.

Previous studies have explored the relationship of MTR analysis to specific proteins, particularly with respect to the use of segmental intolerance analysis to predict or illuminate the linkage of proteins to human disease. This highlights the fact that proteins containing intolerant segments are sometimes subjected to known disease mutations in other parts of the protein and, more rarely, even within the intolerant segment. The latter instance occurs when a disease mutation is observed in a patient-derived database such as ClinVar that is too rare to be seen in the sample of the global (mostly healthy) human population represented by the current gnomAD collection of sequences.

Our objective in this paper differs from the previous work. Here we sought to systematically identify all protein-encoding human genes that contain one or more MTR = 0 (“zero-tolerance”) segments. The resulting list is the main deliverable of this paper. It is hoped that this list will serve as a useful resource for the research community in identifying proteins that contain segments in which mutations result in such catastrophic consequences that they are filtered out of the human population. Evidently, these proteins are profoundly important and/or perilous, such that their study in some cases may yield groundbreaking insight into human biology and molecular pathophysiology. We also reported a few selected observations that can be made regarding the 257 proteins that contain at least one zero-tolerance segment. One important finding is that proteins involved in RNA splicing are the most common group of proteins that contain absolutely intolerant segments. Another important finding is that there are many proteins that contain at least one intolerant segment, but for which there exist no known disease or ClinVar pathogenic mutations to date. We hypothesized that some of these proteins must be essential to human reproduction and/or development, despite, in many cases, having escaped much prior attention or recognition.

2 | RESULTS AND DISCUSSION

2.1 | Human proteins with zero-tolerance segments

We observed 257 proteins—ca. 1.3% of all human proteins—that contain at least one amino acid segment at least 31 residues long (or an N- or C-terminal segment at least 16 residues long) in which amino acid variations appear not to be evolutionarily tolerated (MTR score = 0), as determined by MTR analysis of the human gene sequences in the gnomAD database. These proteins are listed in Table 1, ordered alphabetically by their corresponding gene symbol. For each entry, a variety of supporting information is included, such as the location of the intolerant segment(s) in the protein, the function of the protein, and whether it is a membrane protein. Many of the proteins contain multiple zero-tolerance segments and some of these segments extend well beyond 31 residues. Figure 1a shows a histogram that summarizes the distribution of all possible 31 amino acids segment MTR scores within the 257 proteins. Even within these proteins, only 1.8% of all segments exhibited an MTR score at or near zero. Figure 1b shows the distribution of the whole-protein median MTR score for each protein, where it is seen that the level of genetic tolerance within these proteins is typically not low, with a median score of 0.75 and a mean of 0.71 ± 0.26. These data complement results reported in a column of Table 1 in which the
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|--------------------------------------|----------------|----------|-----------------------|----------------|----------------|------------------------------------------|--------------------------------------|-------------------------------------------|------------------------------------------|
| ABL1 P00519-1 ENST00000318586       | Tyrosine-protein kinase ABL1 | NR tyrosine kinase that is linked to cell growth and survival, as well as chromatin remodeling. Regulates CDC42 signal transduction. | GO:00009790, embryo development | 1,130 | No | 393–423<sup>c</sup> | 0.821 | None | Many |
| ACTB P60709-1<sup>-1</sup> ENST00001675525 | Actin, cytoplasmic 1 | Actin component | GO:0030029, actin filament-based process | 375 | No | 53–91, 124–185, 247–278<sup>1</sup>, (based on ENST00000831789) | 0.1245 | 9 | Many |
| ACTC1 P60032-1 ENST00000290378 | Actin, alpha cardiac muscle 1 | Actin | GO:0060048, cardiac muscle contraction | 377 | No | 105–151 | 0.435 | 4 | Many |
| ACTL6B O94805-1 ENST00000160382 | Actin-like protein 6B | Transcriptional activation and repression of select genes by chromatin remodeling. Role in neuronal development. | GO:0016573, histone acetylation | 426 | No | 1–19 | 0.726 | None | 2 |
| ACTR2 P61360-1 ENST00000260641 | Actin-related protein 2 | ATP binding component of Arp23 complex. | GO:00007010, cytoskeleton organization | 394 | No | 1–16 | 0.654 | None | None |
| AGO2 Q9UKV8-1 ENST00000220382 | Protein argonaute-2 | Essential for RNAi. May inhibit translation. | GO:1905114, cell surface receptor signaling pathway involved in cell-cell signaling | 859 | No | 446–485 | 0.554 | None | None |
| AP2M1 Q6CW1-1 ENST00000292807 | AP-2 complex subunit mu | Component of AP-2. Adaptor protein that plays a role in trafficking. | GO:1905114, cell surface receptor signaling pathway involved in cell-cell signaling | 435 | No | 403–447<sup>c</sup> | 0.4645 | None | None |
| AR P10275-1 ENST00000374690 | Androgen receptor | Steroid hormone receptor that can affect proliferation and differentiation. | GO:00009790, embryo development | 920 | No | 891–932 | 0.849 | None | None |
| ARF1 P84077-1 ENST00000541382 | ADP-ribosylation factor 1 | GTP binding protein involved in protein trafficking | GO:0032880, regulation of protein localization | 181 | No | 16–59 | 0.448 | 1 | 3 |
| ARF5 P84085-1 ENST0000002333 | ADP-ribosylation factor 5 | GTP-binding protein involved in protein trafficking | GO:0006886, intracellular protein transport | 180 | No | 41–74 | 0.603 | None | None |
| ARIH1 Q9Y4X3-1 ENST00000379887 | E3 ubiquitin-protein ligase ARIH1 | E3 ubiquitin ligase. Interacts with cullin-RING ubiquitin ligase complexes. | GO:0000209, protein polyubiquitination | 557 | No | 336–369, 450–483 | 0.5155 | None | None |
| ATF2 P15336-1 ENST00000264110 | Cyclic AMP-dependent transcription factor ATF-2 | Transcriptional activator that involves anti-apoptosis, cell growth, and DNA damage response. Can impair mitochondrial membrane potential. | GO:0045930, negative regulation of mitotic cell cycle | 505 | No | 365–385 | 0.848 | None | 2 |

(Continues)
| Gene symbol, UniProt ID, transcript ID | Encoded protein                                                                 | Function                                                                 | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|---------------------------------------|----------------------------------------------------------------------------------|---------------------------------------------------------------------------|-----------------------|----------------|--------------|------------------------------------------|---------------------------------------|---------------------------------------------|-------------------------------------------|
| ATP1A1 P05023-1 ENST00000295398       | Sodium/potassium-transporting ATPase subunit alpha-1                            | Sodium potassium pump                                                     | GO:0030001; metal ion transport | 1,023          | Yes          | 604–637                                  | 0.537                                 | None                                        | 8                                         |
| ATP1A3 PI3637-2d ENST00000543770      | Sodium/potassium-transporting ATPase subunit alpha-3                            | Sodium potassium pump                                                     | GO:0030001; metal ion transport | 1,024          | Yes          | 355–388d                                 | 0.494                                 | 5                                           | Many                                      |
| ATP2B1 P0020-3 ENST00000428670        | Plasma membrane calcium-transporting ATPase 1                                   | Calcium transporter                                                       | GO:0030001; metal ion transport | 1,220          | Yes          | 421–451                                  | 0.712                                 | None                                        | None                                      |
| ATP9A0C P27449-1 ENST00000330398      | V-type proton ATPase 16 kDa protolipid subunit                                 | Proton-conducting pore forming subunit of the membrane integral V0 complex of vacuolar ATPase responsible for acidifying a variety of intracellular compartments in eukaryotic cells. | GO:0030001; metal ion transport | 155            | Yes          | 133–170                                  | 0.3715                                | None                                        | None                                      |
| ATRX P46300-1 ENST0000373344          | Transcriptional regulator ATRX                                                 | Involved in transcriptional regulation and chromatin remodeling. May be involved in telomere maintenance. | GO:00065004; protein-DNA complex assembly | 2,492          | No           | 1,782–1,814, 2,095–2,142, 2,159–2,213 | 0.837                                 | 3                                           | Many                                      |
| BCL11B Q9C0K0-1 ENST0000357195        | B-cell lymphoma/leukemia 11B                                                   | Key regulator of differentiation and survival of T-lymphocytes. Required for CCK7 and CCR9 receptors. | GO:0000904; cell differentiation | 894            | No           | 789–822                                  | 0.675                                 | 1                                           | 4                                         |
| BRD4 O60855-1 ENST0000263377          | Bromodomain-containing protein 4                                                | Chromatin reader protein that binds acetylated histones and plays a role in epigenetics. | GO:0031056; regulation of histone modification | 1,362          | No           | 508–542                                  | 0.76                                  | None                                        | 1                                         |
| BRD8 Q9H10E9-1 ENST00000254900        | Bromodomain-containing protein 8                                                | May act as a coactivator during transcriptional activation by hormone-activated nuclear receptors. Component of NuA4 histone acetyltransferase. | GO:0016573; histone acetyltransferase | 1,235          | No           | 704–736                                   | 0.891                                 | None                                        | None                                      |
| CACNA1A Q00555-8 ENST0000360228        | Voltage-dependent P/Q-type calcium channel subunit alpha-1A                    | Voltage dependent calcium channel                                         | GO:0030001; metal ion transport | 2,506          | Yes          | 287–325                                  | 0.762                                 | 1                                           | Many                                      |
| CACNA1C Q13936-11a ENST0000347398      | Voltage-dependent L-type calcium channel subunit alpha-1C                      | Calcium channel                                                           | GO:0030001; metal ion transport | 2,186          | Yes          | 731–764d                                 | 0.710                                 | 1                                           | Many                                      |
| CACNA1E Q15878-1 ENST0000367573        | Voltage-dependent R-type calcium channel subunit alpha-1E                      | Voltage gated calcium channel                                             | GO:0030001; metal ion transport | 2,313          | Yes          | 1,648–1,679                              | 0.786                                 | None                                        | 17                                        |
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|-------------------------------------|----------------|----------|-----------------------|----------------|----------------|---------------------|-----------------------------------|-----------------------------------------------|-----------------------------------------------|
| CALM1 P0DP23-1 ENST00000356978     | Calmodulin-1   | Modulates the function of numerous proteins in a calcium dependent manner. Involved in centrosome cycle and cytokinesis. | GO:1905114, cell surface receptor signaling pathway involved in cell-cell signaling | 149            | No             | 110–134             | 0.3975                           | None                                         | 12                                           |
| CALM2 P0DP24-1 ENST00000272288     | Calmodulin-2   | Controls a large number of enzymes and, with CCP110 and centrin, is involved in the centrosome cycle and progression through cytokinesis. | GO:0055074, calcium ion homeostasis | 149            | No             | 78–118              | 0.4675                           | 3                                            | 12                                           |
| CAMK2A Q9UQM7-1 ENST00000348628    | Calcium/calmodulin-dependent protein kinase type II subunit alpha | Kinase that is activated by calcium or calmodulin | GO:0030001, metal ion transport; GO:1905114 | 478            | No             | 111–163             | 0.515                            | None                                         | 9                                            |
| CAND1 Q66VP5-1 ENST00000545506     | Culin-associated NEDD8-associated protein 1 | Key assembly factor of SCF ubiquitin ligase | GO:0010265, SCF complex assembly | 1,230          |                | 46–76               | 0.786                            | None                                         | 1                                            |
| CASK O44936-1 ENST00000378386      | Peripheral plasma membrane protein CASK | Neuronal development protein trafficking | NO:0030002, metal ion transport | 926            |                | 73–103              | 0.652                            | None                                         | Many                                         |
| CDC42 P60953-2 ENST00000400259     | Cell division control protein 42 homolog | Epithelial polarization, attachment of spindle to microtubules. Cell migration. Present in neuronal cells. | GO:1905114, cell surface receptor signaling pathway involved in cell-cell signaling | 191            | No             | 28–109              | 0.2035                           | 5                                            | 11                                           |
| CDC73 Q6TJ9-1 ENST0000036745      | Parallibomin   | RNA pol II recruitment (PAF interaction). Recruits E2 ligases to histones. | GO:0050684, regulation of mRNA processing; GO:1905114 | 531            | No             | 133–173             | 0.721                            | None                                         | Many                                         |
| CDK11B P21127-1 ENST00000407349    | Cyclin-dependent kinase 1B | Cyclin dependent kinase involved in many roles. PremRNA splicing. | GO:0050684, regulation of mRNA processing | 795            | No             | 733–801             | 0.7135                           | None                                         | None                                         |
| CELF2 Q93519-1 ENST00000416582     | CUGBP Elav-like family member 2 | RNA splicing | GO:0050684, regulation of mRNA processing | 508            | No             | 413–449             | 0.552                            | None                                         | None                                         |
| CHD2 O14647-1 ENST00000394996      | Chromodomain-helicase-DNA-binding protein 2 | DNA binding helicase. Promotes deposition of histone H3.3. | GO:0032508, DNA duplex unwinding | 1,828          | No             | 484–519             | 0.743                            | 1                                            | Many                                         |
| CHD4 Q14839-1 ENST0000054406       | Chromodomain-helicase-DNA-binding protein 4 | Part of NuRD complex and remodels chromatin | GO:0043044, ATP-dependent chromatin remodeling | 1,912          | No             | 1,110–1,160, 1,165–1,212 | 0.672                           | 2                                            | 17                                           |
| CLASRP Q8N2M8-1 ENST00000391953    | CLK4-associating serine/arginine rich protein | Probably functions as an alternative splice regulator. | GO:0008380, RNA splicing | 674            | No             | 1–32                | 0.8205                           | None                                         | None                                         |

(Continues)
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|--------------------------------------|-----------------|----------|----------------------|----------------|---------------|------------------------------------------|--------------------------------------|---------------------------------|----------------------------------|
| CLCN4 P13793-1 ENST00000380833       | H+ / Cl(−) exchange transporter 4 | Hydrogen chloride outward rectifying exchanger | GO:0006811, ion transport | 760            | Yes           | 519–549                                  | 0.584                                | 1                              | Many                             |
| CLTC Q00510-1 ENST00002632282        | Clathrin heavy chain 1 | Central protein of clathrin coated pits. Key role in endocytosis. | GO:0030001, metal ion transport | 1,675          | No            | 1,302–1,336                              | 0.660                                | None                           | 5                                |
| CNOT6L Q06551-1 ENST0000050423       | CCR4-NOT transcription complex subunit 1-like | Has poly(A) exonuclease activity. Catalytic component of the CCR4-NOT complex. | GO:00066402, mRNA catabolic process | 555            | No            | 404–454                                  | 0.7255                               | None                           | None                            |
| CPSF4 O96549-1 ENST0000029247       | Cleavage and polyadenylation specificity factor subunit 4 | Pre-mRNA processing. Poly-A cap | GO:0006684, regulation of mRNA processing | 269            | No            | 68–115                                   | 0.585                                | None                           | None                            |
| CREB1 P1620-2 ENST0000043024        | Cyclic AMP-responsive element-binding protein 1 | Phosphorylation-dependent transcription factor. Binds to CRE and is enhanced by TORC coactivators. Circadian rhythm and differentiation of adipose tissue. | GO:0007623, circadian rhythm | 327            | No            | 271–315c                                  | 0.668                                | None                           | 1                                |
| CREBL2 O60519-1 ENST0000228865       | CAMP-responsive element-binding protein-like 2 | May play a role in cell cycle. Transcriptional activity involved in adipose differentiation. | GO:0006351, transcription, DNA-templated | 120            | No            | 20–54                                    | 0.836                                | None                           | None                            |
| CSNK2B P67780-1 ENST0000375082       | Casein kinase II subunit beta | Regulatory subunit of casein kinase 2, a normally constitutively active kinase. Participates in Wnt signaling. | GO:1905114, cell surface receptor signaling pathway involved in cell–cell signaling | 215            | No            | 1–19                                     | 0.5695                               | 1                              | 3                                |
| CSTF2 P13340-1 ENST00000372972       | Cleavage stimulation factor subunit 2 | Required for polyadenylation and pre-mRNA cleavage | GO:0006379, mRNA cleavage | 577            | No            | 555–577c                                  | 0.8155                               | None                           | None                            |
| CTCF P4971-1 ENST0000026480          | Transcriptional repressor CTCF | Involved in transcriptional regulation by binding to chromatin insulators. Plays a role in CENPE recruitment during mitosis. | GO:0071824, protein-DNA complex subunit organization | 727            | No            | 279–324                                  | 0.6235                               | None                           | 16                              |
| CUL1 Q15161-1 ENST00003255222        | Culin-1 | Core component of cullin-RING-based SCF E3 ubiquitin ligase in ubiquitination of proteins involved in cell cycle progression. | GO:1905114, cell surface receptor signaling pathway involved in cell–cell signaling | 776            | No            | 532–566                                  | 0.484                                | None                           | None                            |
| CUL4B Q15162-1 ENST00000371322       | Culin-4B | Core component of cullin-RING-based E3 ubiquitin ligase | GO:00016687, protein ubiquitination | 913            | No            | 709–742c                                  | 0.631                                | None                           | Many                            |
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? (UniProt numbering) | Intolerant segment(s) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|--------------------------------------|-----------------|----------|-----------------------|----------------|-------------------------------------|----------------------|-----------------------------------|-----------------------------------------------|-----------------------------------------------|
| DDX3X O00571-2, ENST00000457138     | ATP-dependent RNA helicase | ATP-dependent RNA helicase. Binds RNA G4s. Transcription regulation. Required for ATF4 mRNA translation. Mediates virus replication. | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 646            | No                                  | 476–507             | 0.603                            | 1                                             | Many                                           |
| DENND1A Q0TEH3-1, ENST00000373624  | DENN domain-containing protein 1A | Guanine nucleotide exchange factor regulating clathrin endocytosis through RAB35 activation | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 1,009          | No                                  | 1–18                             | 0.875                            | None                                          | None                                           |
| DHX15 O43143-1, ENST0000036812     | Pre-mRNA-splicing factor ATP-dependent RNA helicase | Pre-mRNA processing factor involved in disassembly of spliceosomes. | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 795            | No                                  | 462–509, 532–573               | 0.501                            | None                                          | None                                           |
| DHX9 Q08211-1, ENST0000036749      | ATP-dependent RNA helicase A | Helicase activity. Some mRNA splicing activity. | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 1,270          | No                                  | 708–757                         | 0.666                            | None                                          | None                                           |
| DKC1 O60832-1, ENST00000369350     | H/ACA ribonucleoprotein complex subunit DKC1 | Catalyses uridine to pseudouridine in RNA | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 514            | No                                  | 88–138, 167–207, 218–236, 371–404 | 0.584                            | 2                                             | Many                                           |
| DLS3 Q2796-1, ENST0000037480       | Disks large homolog 3 | Role in learning, through NMDA receptor signaling | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 817            | No                                  | 522–572                         | 0.800                            | None                                          | 14                                             |
| DUSP8 Q3320-1, ENST0000039724      | Dual specificity protein phosphatase 8 | Phosphatase that regulates MAPK activity | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 625            | No                                  | 610–625                         | 0.780                            | None                                          | None                                           |
| EBP1 Q8ND8-1, ENST00000263991      | EH domain-binding protein 1 | May play a role in actin reorganization. | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 1,231          | No                                  | 1–21                            | 0.931                            | None                                          | None                                           |
| EHMT2 Q66KJ7-1, ENST0000037537      | Histone-lysine N-methyltransferase EHMT2 | Histone methyltransferase that mono or di-methylates H3K9 | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 1,210          | No                                  | 1,070–1,108                    | 0.823                            | None                                          | None                                           |
| EIF1AX P4781-1, ENST0000037907      | Eukaryotic translation initiation factor 1A, X-chromosomal | Seems to be required for maximal protein biosynthesis. | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 144            | No                                  | 5–45, 56–128                    | 0.2105                           | None                                          | None                                           |
| EIF1AY Q4602-1, ENST0000036135      | Eukaryotic translation initiation factor 1A, Y-chromosomal | Seems to be required for maximal protein biosynthesis. | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 144            | No                                  | 124–344                         | 0.5075                           | None                                          | None                                           |
| EIF2S2 P20042-1, ENST0000037480     | Eukaryotic translation initiation factor 2 subunit 2 | Initiation of translation | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 333            | No                                  | 226–285, 318–333                | 0.645                            | None                                          | None                                           |
| EIF2S3 P41091-1, ENST00000253089    | Eukaryotic translation initiation factor 2 subunit 3 | Subunit of eIF-2 involved in early steps of protein synthesis. | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 472            | No                                  | 150–204, 429–461                 | 0.478                            | None                                          | 6                                               |
| Gene symbol | UniProt ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|-------------|------------|-----------------|----------|-----------------------|----------------|---------------|----------------------|--------------------------------|------------------------------------------|-----------------------------------------------|
| EIF3A       | Q14152-1   | EIF3A           | Eukaryotic translation initiation factor 3 subunit A | Subunit of the eIF-3 complex. Required for protein synthesis. Targets a subset of mRNA involved in cell proliferation. | GO:0006413, translation initiation | 1,382 | No | 1–18 | 0.887 | None | None |
| EIF4A3      | P38919-1   | EIF4A3          | Eukaryotic initiation factor 4A-III | ATP dependent helicase. PremRNA splicing. Core component of exon junction complex. Involved in craniosacral development. | GO:0009790, embryo development | 411 | No | 204–254 (based on ENST000000649764) | 0.4735 | None | 1 |
| ERH         | P44080-1   | ERH             | Enhancer of rudimentary homolog | May have a role in cell cycle | GO:0007049, cell cycle | 104 | No | 30–61 | 0.325 | None | None |
| EEF1        | P62495-1   | EEF1            | Eukaryotic peptide chain release factor subunit 1 | Directs termination of nascent peptide synthesis in response to stop codons. Component of SURF complex. | GO:0002184, cytoplasmic translational termination | 437 | No | 55–87, 324–355 | 0.490 | None | None |
| F8          | P00451-1   | F8              | Coagulation factor VIII | Factor VIII, along with calcium and phospholipid, acts as a co-factor for F9/factor IXa, when it converts F10/factor X to the activated form, factor Xa. | GO:00016491, oxidoreductase activity | 2,351 | No | 95–131 | 0.893 | 4 | Many |
| FGU1        | P91374-1   | FGU1            | FYVE, RhoGEF, and PH domain-containing protein 1 | Activates CDC42. Plays a role in cytoskeleton and cell shape | GO:0007010, cytoskeleton organization | 961 | No | 575–616, 739–769 | 0.7615 | 1 | Many |
| FMRI1       | Q06787-1   | FMRI1           | Synaptic functional regulator | mRNA regulation. Maybe DNA repair in neuronal cells. | GO:00050684, regulation of mRNA processing | 632 | No | 69–99 | 0.858 | None | 16 |
| FOXG1       | P53186-1   | FOXG1           | Forkhead box protein G1 | Transcription repression factor important for neurogenesis. | GO:0007420, brain development | 489 | No | 175–209, 217–247 | 0.564 | 2 | Many |
| FOXJ3       | Q9U790-1   | FOXJ3           | Forkhead box protein J3 | Transcriptional activator of MEF2C. Plays an important role in spermatogenesis. | GO:00010468, regulation of gene expression | 622 | No | 100–139 | 0.848 | None | None |
| GABPA       | Q06546-1   | GABPA           | GABP-binding protein alpha chain | Transcription factor capable of interacting with purine rich repeats. | GO:0009790, embryo development | 454 | No | 376–406 | 0.698 | None | None |
| GABRA2      | P34903-1   | GABRA2          | GABA receptor subunit alpha-2 | Ligand gated chloride channel that is a component of the receptor for GABA. | GO:00099536, synaptic signaling | 451 | Yes | 279–316 | 0.6545 | None | None |
| GABRA3      | P34903-1   | GABRA3          | GABA receptor subunit alpha-3 | GABA receptor | GO:00099536, synaptic signaling | 492 | Yes | 306–338 | 0.688 | None | None |
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|---------------------------------------|----------------|----------|----------------------|----------------|--------------|------------------------------------------|----------------------------------|-------------------------------------------|----------------------------------------|
| GABRB2 P03137-1 ENST00000274347      | Gamma-aminobutyric acid receptor subunit beta-2 | Ligand-gated chloride channel component of the GABA receptor. | GO:00099536, synaptic signaling | 512            | Yes          | 151–190                                  | 0.618                           | None                                      | Many                                   |
| GDF1 P03137-1 ENST00000274347        | Growth/differentiation factor 11 | Secreted signal involved in development. | GO:0045664, regulation of neuron differentiation | 407            | No           | 1–21                                     | 0.728                           | None                                      | None                                   |
| GJB1 P05004-1 ENST00000374029        | Gap junction beta-1 protein | Forms gap junctions | GO:0007267, cell-cell signaling | 283            | Yes          | 50–89                                    | 0.713                           | 3                                         | Many                                   |
| GLRA2 P23461-1 ENST00000218075       | Glycine receptor subunit alpha-2 | Glycine ligand gated chloride channel. Also triggered by taurine and beta-alanine. | GO:1905114, cell surface receptor signaling pathway involved in cell-cell signaling | 452            | Yes          | 261–315                                  | 0.698                           | None                                      | None                                   |
| GNAL P38405-1 ENST0000053321         | Guanine nucleotide-binding protein G(o) subunit alpha | G protein that may be involved in olfactory and visual transduction. | GO:0019932, second-messenger-mediated signaling | 381            | No           | 35–66, 140–171, 178–215 c | 0.615                           | 1                                         | 6                                      |
| GNAQ P50484-1 ENST00000286548        | Guanine nucleotide-binding protein G(q) subunit alpha | G protein involved in many transmembrane signaling pathways. Is important for B cell selection and chemotaxis of neutrophils and dendritic cells. | GO:0019932, second-messenger-mediated signaling | 359            | No           | 171–202                                  | 0.575                           | None                                      | 4                                      |
| GNAS Q5JSF2-1 ENST0000037100         | Guanine nucleotide-binding protein G(q) subunit alpha isoforms short | G protein that is activated by GPCRs including beta-adrenergic receptors, stimulates Ras signaling. | NA              | 1.037         | No           | 899–959                                  | 0.8455                          | None                                      | 17                                     |
| GNG1 P63215-1 ENST00000294117        | Guanine nucleotide-binding protein G(1)/G(S)/G(O) subunit gamma-3 | G protein subunit and required for GTPase activity. | GO:00055074, calcium ion homeostasis | 75             | No           | 58–75                                    | 0.819                           | None                                      | 1                                      |
| GOLGA8G Q08AF8-1 ENST0000052669       | Putative golgin subfamily A member 8F/8G | Possibly a pseudogene | GO:0000226, microtubule cytoskeleton organization | 430            | No           | 389–422 c (based on ENST00000825590) | 0.9045                          | None                                      | None                                   |
| GORASP2 PQ188Y-1 ENST00000234860      | Golgi reassembly-stacking protein 2 | Role in assembly and membrane stacking of the Golgi cisternae. May regulate intracellular transport. Required for normal acrosome formation in spermiogenesis. Mediates ER stress and induced unconventional trafficking of core-glycosylated CFTR to cell membrane. | GO:0045184, establishment of protein localization | 454            | No           | 1–16                                     | 0.896                           | None                                      | None                                   |

(Continues)
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|--------------------------------------|-----------------|----------|-----------------------|----------------|---------------|------------------------------------------|-----------------------------------|---------------------------------|----------------------------------|
| GRIA2 P42262-1 ENST00000264426      | Glutamate receptor 2 | Receptor for glutamate that functions as an ion channel in the CNS. | GO:0099536, synaptic signaling | 833 | Yes | 521–554 | 0.6745 | None | 1 |
| GRIA2 P42263-1 ENST0000022768       | Glutamate receptor 3 | Glutamate-gated ion channel | GO:0007215, glutamate receptor signaling pathway | 894 | Yes | 487–543, 602–658, 745–777, 879–894 (based on ENST00000371256) | 0.608 | 2 | Many |
| GRIN1 Q5586-1 ENST00000371561       | Glutamate receptor ionotropic, NMDA 1 | NMDA subunit that binds glutamate | GO:0050684, regulation of mRNA processing | 938 | Yes | 549–581, 602–638, 745–777, 879–894 (based on ENST00000371256) | 0.544 | 3 | Many |
| GRIN2A Q12879-1 ENST00000396573     | Glutamate receptor ionotropic, NMDA 2A | Ligand-gated ion channel | GO:0003001, metal ion transport | 1,464 | Yes | 631–670 | 0.799 | 5 | Many |
| GRIN2B Q1324-1 ENST0000060986       | Glutamate receptor ionotropic, NMDA 2B | Component of NMDA receptor complex | GO:1905114, cell surface receptor signaling pathway involved in cell–cell signaling | 1,484 | Yes | 504–567, 662–700, 744–774 | 0.659 | 5 | Many |
| GSPT1 P15570-1 ENST00000563408      | Eukaryotic peptide chain release factor GTP-binding subunit ERF3A | Translation termination | GO:00002184, cytoplasmic translational termination | 499 | No | 170–204 | 0.739 | None | None |
| HCP1 Q15160-1 ENST00000310441       | Host cell factor 1 | Control of cell cycle from G1 to S, Coactivator of GABP2. | GO:00099790, embryonic development | 2035 | No | 143–173, 207–240, 1,600–1,631, 1,976–2,007 | 0.648 | 1 | Many |
| HMGN4 Q00479-1 ENST0000373557       | High mobility group nucleosome-binding domain-containing protein 4 | Chromatin binding | GO:00031492, nucleosomal DNA binding | 90 | No | 1–17 | 0.978 | None | None |
| HNF1B P35680-1 ENST0000017811       | Hepatocyte nuclear factor 1-beta | Transcription factor. Binds to FPC element in PLAU gene. Organ development. | GO:00099790, embryonic development | 557 | No | 1–17 | 0.8195 | None | Many |
| HNRNPC P07910-1 ENST0000054455      | Heterogeneous nuclear ribonucleoproteins C1/C2 | Binds pre-mRNA and nucleates the assembly of 40S hnRNP particles. May play a role in spliceosome assembly and pre-mRNA splicing. | GO:0004387, regulation of RNA stability | 306 | No | 1–17 | 0.67 | None | None |
| HNRNPD Q4103-1 ENST00000313899      | Heterogeneous nuclear ribonucleoprotein D0 | Binds with high affinity to RNA with AU-rich elements. Functions as transcription factor. | GO:00006401, RNA catabolic process | 355 | No | 108–148 | 0.924 | None | None |
| HNRNPII2 P55975-1 ENST00000316594   | Heterogeneous nuclear ribonucleoprotein H2 | Component of hnRNP which processes pre-mRNAs. | GO:0008380, RNA splicing | 449 | No | 78–111, 151–190 | 0.504 | None | 5 |
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|---------------------------------------|-----------------|----------|----------------------|----------------|---------------|-------------------------------------------|-----------------------------------|---------------------------------|---------------------------------|
| HNRNPK P61978-2 ENST00000351839       | Heterogeneous nuclear ribonucleoprotein K | mRNA processing (one of major pre-mRNA binding proteins), DNA binding. TP53 coactivator. | GO:0025659: regulation of mRNA processing | 463            | No            | 441–463 (based on ENST00000876263)        | 0.6985                           | None                           | 8                              |
| HSD17B10 Q99714-1 ENST00000168216     | 3-hydroxyacyl-CoA dehydrogenase type-2 | Mitochondrial dehydrogenase involved in pathways of fatty acid, branched-chain amino acid and steroid metabolism. | GO:001575: organic substance catabolic process | 261            | No            | 145–181                                   | 0.628                            | None                           | 15                             |
| HUWE1 Q7Z6Z7-1 ENST00000342360        | E3 ubiquitin-protein ligase HUWE1 | E3 ubiquitin ligase | GO:0000209: protein polyubiquitination | 4,374          | No            | 499–529, 547–578, 3,006–3,042, 3,917–3,952, 4,385–4,390 | 0.728                            | 1                              | Many                           |
| INT56 Q9UL0-1 ENST00000311234         | Integrator complex subunit 6 | Component of integrator complex. Involved in U1 and U2 transcription. | GO:0006366: transcription by RNA polymerase II | 887            | No            | 76–107                                    | 0.7715                           | None                           | None                           |
| IRAK1 P51471-1 ENST00000369090        | Interleukin-1 receptor-associated kinase 1 | Serine/threonine kinase that plays a critical role in initiating the innate immune system. | GO:0002218: activation of innate immune response | 712            | No            | 26–58                                     | 0.901                           | None                           | 1                              |
| KAT7 O9525-1 ENST00000259021           | Histone acetyltransferase KAT7 | Catalytic component of H4 histone acetyltransferase complexes. | GO:0006573: histone acetylation | 611            | No            | 405–443                                   | 0.678                            | None                           | None                           |
| KCNA3 P22001-1 ENST00000369090        | Potassium voltage-gated channel subfamily A member 3 | Voltage-gated potassium channel | GO:0030001: metal ion transport | 575            | Yes           | 359–392                                   | 0.865                            | None                           | 1                              |
| KCNB1 Q14721-1 ENST00000317134        | Potassium voltage-gated channel subfamily B member 1 | Voltage-gated potassium channels that can form heterotetrameric channels with other potassium channels. | GO:0030001: metal ion transport | 858            | Yes           | 82–113, 326–357, 349–413                   | 0.664                            | 3                              | Many                           |
| KCNC2 Q66PR1-1 ENST00000549446         | Potassium voltage-gated channel subfamily C member 2 | Voltage-gated potassium channel. Also acts in various signaling pathways such as NO signaling. | GO:0030001: metal ion transport | 638            | Yes           | 370–401                                   | 0.720                            | None                           | None                           |
| KCND3 Q9UK71-1 ENST00000315987         | Potassium voltage-gated channel subfamily D member 3 | Voltage-gated inactivated A-type potassium channel. May contribute to current in heart or neuron. | GO:0030001: metal ion transport | 655            | Yes           | 298–332, 364–407                           | 0.7405                           | 3                              | Many                           |
| KCHH7 Q9NS40-1 ENST0000023342          | Potassium voltage-gated channel subfamily H member 7 | Voltage-gated potassium channel | GO:0030001: metal ion transport | 1,196          | Yes           | 612–642                                   | 0.834                            | None                           | 1                              |
| KCNJ3 P48549-1 ENST00000295001         | G protein-activated inward rectifier potassium channel 1 | Inward rectifier potassium channel controlled by G protein and playing a crucial role in regulating heartbeat. | GO:0030001: metal ion transport | 501            | Yes           | 161–194                                   | 0.6365                           | None                           | None                           |

(Continues)
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|----------------------------------------|-----------------|----------|----------------------|----------------|---------------|------------------------------------------|-------------------------------|---------------------------------|---------------------------------|
| KCNA1 Q12791-1 ENST00000286628         | Calcium-activated potassium channel subunit alpha-1 | Export of potassium triggered by changes in cytosolic calcium or magnesium. Regulates smooth muscles, hair cells in cochlea, transmitter release, and innate immunity. | GO:00030001; metal ion transport | 1,236 | Yes | 554–988, 1,009–1,039 | 0.681 | None | Many |
| KCNQ2 O43526-1 ENST0000359325          | Potassium voltage-gated channel subfamily KQT member 2 | Heterotetramers with KCNQ3 to form a voltage gated channel important for regulation of neuronal excitability. | GO:00030001; metal ion transport | 872 | Yes | 197–237 | 0.721 | 5 | Many |
| KDM2A Q9YJK7-1 ENST00000529006         | Lysine-specific demethylase 2A | Histone demethylase that preferentially demethylates H3K36. Regulates circadian clock. | GO:0007623; circadian rhythm | 1,162 | No | 275–311, 585–627 | 0.704 | None | None |
| KDM3B Q7LBC6-1 ENST00000314358         | Lysine-specific demethylase 3B | Histone demethylase that specifically demethylates histone H3. | GO:00016570; histone modification | 1,761 | No | 1,678–1714, 1716–1748 | 0.786 | None | None |
| KIF11 P52732-1 ENST00000260731          | Kinesin-like protein KIF11 | Motor protein required for establishing a bipolar spindle during mitosis. Also involved in Golgi-to-cell surface trafficking. | GO:00007623; circadian rhythm | 1,056 | No | 259–309 | 0.840 | None | 13 |
| KIF1A Q12796-1 ENST00000498729         | Kinesin-like protein KIF1A | Motor for anterograde axonal transport of synaptic vesicle precursors. Interacts with CALM1. Required for neuronal dense core vesicles transport to dendritic spines and axons. | GO:00030705; cytoskeleton-dependent intracellular transport | 1,791 | No | 1,465–1,498 | 0.779 | 1 | Many |
| KIF5A Q12840-1 ENST00000455537          | Kinesin heavy chain isoform 5A | Kinesin transport of neurofilament proteins | GO:00030705; cytoskeleton-dependent intracellular transport | 1,032 | No | 230–264 | 0.788 | 1 | Many |
| KMT2C Q8NEZ4-1 ENST00000262389          | Histone-lysine N-methyltransferase 2C | Histone methyltransferase to H3K4. Chromatin remodeling. | GO:00016571; histone methylation | 4,911 | No | 349–379 | 0.923 | None | Many |
| KPNB1 Q14974-1 ENST00000290358          | Importin subunit beta-1 | Binds to nuclear localization signals and imports proteins into the nucleus. | GO:0005169; nuclear transport | 876 | No | 706–715 | 0.580 | None | None |
| LPA P08519-1 ENST00000316300            | Apolipoprotein(a) | Main constituent of lipoprotein(a). Serine protease activity. Inhibits plasminogen activator 1. | GO:0006508; proteolysis | 4,584 | No | 99–130 | 0.935 | None | None |
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|--------------------------------------|----------------|----------|-----------------------|----------------|---------------|-----------------------------------------|-----------------------------------|-----------------------------------------------|-----------------------------------------------|
| LUC7L3 O95232-1 ENST00000505685     | Luc7-like protein 3 | Binds cAMP regulatory element DNA sequence. May play a role in RNA splicing | GO:0008380; RNA splicing | 432          | No             | 185–225                                 | 0.908                             | None                                          | None                                          |
| MAML1 Q1349-4 ENST0000042663       | Mastermind-like domain-containing protein 1 | Transactivates HES3 independent of NOTCH | GO:0006357; regulation of transcription by RNA polymerase II | 749          | No             | 713–747 (based on ENST00000432680)    | 0.916                             | None                                          | None                                          |
| MEF2C Q06413-1 ENST00000374086     | Myocyte-specific enhancer factor 2C | Transcription activator that binds specifically to MEF2 element in many muscle-specific genes. Controls cardiac morphogenesis and myogenesis. Plays a role in hippocampal learning. Important for immune cells. | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 2,177        | No             | 1,138–1,169                             | 0.693                             | 1                                             | Many                                          |
| MEDE14 O61244-1 ENST00000324817     | Mediator of RNA polymerase II transcription subunit 12 | Component of mediator complex. Involved in the regulation of nearly all RNA pol-II dependent genes. May specifically regulate transcription of targets of Wnt signaling pathway and SHH signaling. | GO:0006366; transcription by RNA polymerase II | 1,454        | No             | 1,277–1,308                             | 0.845                             | None                                          | None                                          |
| MEF2C Q06413-1 ENST00000374086     | Myocyte-specific enhancer factor 2C | Transcription activator that binds specifically to MEF2 element in many muscle-specific genes. Controls cardiac morphogenesis and myogenesis. Plays a role in hippocampal learning. Important for immune cells. | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 473          | No             | 1–36                                    | 0.6485                            | 6                                             | Many                                          |
| METTL14 Q00138-1 ENST0000038822     | N6-adenosine-methyltransferase non-catalytic subunit | Component of methyltransferase complex that methylates at the N6 position of some mRNAs and regulates circadian rhythm, differentiation of embryonic stem cells and cortical neurogenesis. | GO:00032259; methylation | 456          | No             | 104–135                                 | 0.801                             | None                                          | None                                          |
| MMP16 P5152-1 ENST0000028664        | Matrix metalloproteinase-16 | Endopeptidase that degrades components of extracellular matrix. Matrix remodeling of blood vessels. | GO:0009790; embryo development | 607          | Yes            | 199–233                                 | 0.7975                            | None                                          | None                                          |
| MOB4 Q09933-1 ENST00000323303       | MOB-like protein phos (cont) | May play a role in membrane trafficking, specifically membrane budding. | GO:0046872; metal ion binding | 225          | No             | 123–153                                 | 0.7225                            | None                                          | None                                          |
| Gene symbol, | UniProt ID, | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|-------------|-------------|-----------------|----------|-----------------------|---------------|---------------|----------------------|-----------------------------------|---------------------------------|----------------------------------------|
| MRC1        | P22897-1    | Macrophage mannose receptor 1 | Mediates endocytosis of glycoproteins. | GO:0044419; interspecies interaction between organisms | 1,456 | Yes | 168–213, 411–471 (based on ENST00000239761) | 0.778 | None | 1 |
| MYB         | P1042-1     | Transcriptional activator Myb | Transcriptional activator. DNA-binding to YAAC[GT]G. Plays an important role in the control of proliferation and differentiation of hematopoietic progenitor cells. | GO:00066338; chromatin remodeling | 640 | No | 118–157 | 0.841 | None | None |
| NAA10       | P4127-1     | N-alpha-acetyltransferase 10 | Acetyltransferase, particularly the first amino acid following removal of methionine. | GO:0006473; protein acetylation | 235 | No | 17–47, 53–95, 104–148 | 0.4335 | 6 | None |
| NAA15       | Q9BXJ9-1    | N-alpha-acetyltransferase 15, NatA auxiliary subunit | Auxiliary subunit of N-terminal acetyltransferase activity. May be important for vascular, hematopoietic and neuronal growth and development. Required to control retinal neovascularization. | GO:0006473; protein acetylation | 866 | No | 97–137 | 0.795 | None | None |
| NEDD4       | Q15843-1    | NEDD8          | Plays an important role in cell cycle control and embryogenesis via its conjugation to target proteins. Ubiquitin-like | GO:00043687; post-translational protein modification | 81 | No | 16–46 | 0.4175 | None | None |
| NIPBL       | Q6K79-1     | Nipped-B-like protein | Loading of cohesion complex onto chromatin. | GO:0009790; embryo development | 2,804 | No | 2,073–2,107 | 0.816 | 3 | Many |
| NONO        | QI5233-1    | Non-POU domain-containing octamer-binding protein | Plays a variety of roles in nuclear processes. | GO:0006281; DNA repair | 471 | No | 174–216, 221–265 | 0.600 | None | 5 |
| NRA2A       | P43354-1    | Nuclear receptor subfamily 4 group A member 2 | Transcriptional regulator for differentiation of neurons during development. Crucial for expression of SLC6A3, SLC18A2, TH, and DRD2. | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 598 | No | 261–293, 318–348 | 0.825 | None | None |
| NRBP1       | Q6W7Y1-1    | Nuclear receptor-binding protein | May play a role in trafficking between ER and the Golgi through interaction with rho-type GTPases. | GO:0006810; transport | 535 | No | 199–322 (based on ENST00000379863) | 0.671 | None | None |
| NSMF        | Q6W7Y1-1    | NMDA receptor synaptomodulin signaling and neuronal migration factor | Part of CREB shut off pathway. Couples NMDA-sensitive glutamate receptor and triggers long lasting changes to dendrites and synapses. | GO:0048814; regulation of dendrite morphogenesis | 530 | No | 1–19 | 0.804 | None | None |
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|-------------------------------------|----------------|----------|-----------------------|----------------|----------------|------------------------------------------|----------------------------------|------------------------------------------|------------------------------------------|
| NUDT11 Q96G61-1 ENST00000375992     | Diphosphoinositol polyphosphate phosphohydrolase 3-beta | Clauses a beta-phosphate from the diphosphate groups in PP-InsP5 | GO:0009058; biosynthetic process | 164 | No | 1–20 | 0.522 | None | None |
| NUDT21 O43809-1 ENST0000030291     | Cleavage and polyadenylation specificity factor subunit 5 | Component of cleavage factor Im. Involved in mRNA processing. | GO:0050684; regulation of mRNA processing | 227 | No | 170–215 | 0.5065 | None | None |
| OGT O15294-1 ENST00000373719      | UDP-N-acetylglucosamine-peptide N-acetylglucosaminyltransferase 110 kDa subunit | Glycosylates other proteins. | GO:0006493; protein O-linked glycosylation | 1,046 | Yes | None | 0.542 | None | None |
| OR4F17 Q8NGA8-1 ENST00000585993   | Olfactory receptor 4F17 | Predicted olfactory receptor | GO:0007165; signal transduction | 305 | Yes | 1–22 | 0.938 | None | None |
| OTUD5 Q96734-1 ENST00000156084   | OTU domain-containing protein 5 | Deubiquitinating functioning as a negative regulator of immune system. | GO:00016579; protein deubiquitination | 571 | No | 171–201, 343–411 | 0.5085 | None | 1 |
| PAK2 Q31777-1 ENST00000327334   | Serine/threonine-protein kinase PAK2 | Serine/threonine kinase. Involved in cytoskeleton regulation, cell motility, cell cycle progression apoptosis, or proliferation. Downstream of CDC42 and RAC1. | GO:00031098; stress-activated protein kinase signaling cascade | 524 | No | 362–397 | 0.719 | None | None |
| PAK3 O75914-1 ENST00000372080 | Serine/threonine-protein kinase PAK3 | Serine/threonine kinase that affects cytoskeleton regulation, cell migration, and cell cycle. Acts downstream of CDC42. | GO:0006468; protein phosphorylation | 559 | No | 68–99, 290–335, 413–455, 458–489 | 0.645 | 1 | 15 |
| PBX1 P40424-1 ENST0000042086     | Pre-B-cell leukemia transcription factor 1 | Binds DNA in conjunction with HOX proteins. Spleen development. | GO:0009790; embryo development | 430 | No | 275–306 | 0.632 | None | 4 |
| PCBP2 Q15366-1 ENST00000439980   | Poly(C)-binding protein 2 | Single strand nucleotide binding protein that preferentially binds to C. Acts as adaptor between MAVS and E3 ITCH | GO:00043161; proteasome-mediated ubiquitin-dependent protein catalytic process | 365 | No | 90–120 | 0.4825 | None | None |
| PCIY1B Q7VYK3-1 ENST0000379344   | Choline-phosphate cytidylyltransferase B | Rate-limiting step in the CDP-choline pathway for phosphatidylcholine biosynthesis | GO:00009058; biosynthetic process | 369 | No | 220–268 | 0.675 | None | None |
| PHF5A Q8HTV0-1 ENST00000216252   | PHD finger-like domain-containing protein 5A | Involved in PAF1 complex in transcriptional elongation. Involved in pre-mRNA splicing and deposition of certain histones. | GO:0000397; mRNA processing | 110 | No | 1–18, 85–110 | 0.324 | None | None |
| Gene symbol | UniProt ID | transcript ID | Encoded protein | Function | GO pathway or process | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) |
|-------------|------------|---------------|----------------|----------|----------------------|-----------------------------------|----------------------------------------|-----------------------------------------------|----------------|----------------|------------------------------------------|
| PIK3CA      | P42336-1   | ENST00000263967 | Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform | Subunit of PI3K | GO:0009749; response to glucose | 1,068 | No | 927-971, 1,010-1,041 | 1,068 | No | 975, 1,010-1,041 |
| PLS3        | P13797-1   | ENST00000355899 | Plastin-3 Actin bundling proteins found in microvilli, stereocilia, filopodia and may play a role in bone development. | cytoskeleton organization | GO:0007010 | 630 | No | 450-503 | 630 | No | None None |
| POLR2A      | P24928-1   | NA             | DNA-directed RNA polymerase II subunit RPB1 | Forms RNA polymerase active center with another catalytic subunit. | GO:0006366; transcription by RNA polymerase II | 1,970 | No | 476-506 | 1,970 | No | 506-532, 544-578, 765-775 |
| POLR2B      | P30876-1   | ENST00000381227 | DNA-directed RNA polymerase II subunit RPB2 | DNA dependent RNA polymerase catalyzing transcription. | GO:0006366; transcription by RNA polymerase II | 1,174 | No | 490-522, 524-557, 746-778, 979-1,026, 1,072-1,135 | 1,174 | No | None None |
| POU3F2      | P20265-1   | ENST00000328345 | Histone-lysine N-methyltransferase EHMT2 | Histone methyltransferase that mono or di-methylates Lys-9. | GO:0006479; protein methylation | 443 | No | 270-344 | 443 | No | None None |
| POU3F3      | P20264-1   | ENST00000361360 | POU domain, class 3, transcription factor 3 | Transcription factor that acts synergistically with SOX11 and SOX4 Role in neuronal development. | GO:0030900; forebrain development | 500 | No | 317-352 | 500 | No | None None |
| PPP1CB      | P05141-1   | ENST00000395866 | Serine/threonine-protein phosphatase PIP-beta catalytic subunit | Protein phosphatase that forms complexes with over 200 regulatory proteins. | GO:0006468; protein phosphorylation | 327 | No | 51-113 | 327 | No | None None |
| PPP2CA      | P05141-1   | ENST00000481295 | Serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform | Major phosphatase for microtubule-associated protein. | GO:0006468; protein phosphorylation | 309 | No | 142-381 | 309 | No | None None |
| PPP3R1      | P05141-1   | ENST00000403180 | Calcineurin subunit B type 1 | Regulatory subunit of calcineurin, a calcium-dependent, calmodulin-stimulated protein phosphatase. | GO:1905114; cell surface receptor signaling pathway involved in cell–cell signaling | 170 | No | 65-96 | 170 | No | None None |
| PRPF4B      | P05141-1   | ENST00000423600 | Serine/threonine protein kinase | Protein that interacts with PI3K and PI4K complex. | GO:0006468; protein phosphorylation | 1,007 | No | 811-941 | 1,007 | No | None None |
| PSD4        | P05141-1   | ENST00000437099 | Pre-mRNA processing factor | Pre-mRNA processing, splicing vs. translation | GO:0006468; protein phosphorylation | 2,135 | No | 603-764, 837-879, 851-941, 1,088-1,199 | 2,135 | No | None None |
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|--------------------------------------|----------------|----------|----------------------|----------------|---------------|------------------------------------------|---------------------------------|------------------------------------------|------------------------------------------|
| PRPS1 P60891-1 ENST0000372435        | Ribose-phosphate pyrophosphokinase 1 | Essential for nucleotide synthesis. | GO:0019438, aromatic compound biosynthetic process | 318            | No            | 1–30, 84–121, 123–162, 168–199          | 0.446                          | 7                                        | Many                                     |
| PSMC1 P62191-1 ENST0000261303        | 26S proteasome regulatory subunit 4   | 26S proteasome subunit | GO:1905114, cell surface receptor signaling pathway involved in cell-cell signaling | 440            | No            | 291–332                                  | 0.652                          | None                                     | None                                     |
| PSMC2 P13598-1 ENST0000435765        | 26S proteasome regulatory subunit 7   | Component of 26S proteasome | GO:1905114, cell surface receptor signaling pathway involved in cell-cell signaling | 433            | No            | 284–318                                  | 0.644                          | None                                     | None                                     |
| PSMC5 P62195-1 ENST0000310344        | 26S proteasome regulatory subunit 8   | 26S proteasome subunit | GO:1905114, cell surface receptor signaling pathway involved in cell-cell signaling | 406            | No            | 112–144, 158–188                         | 0.563                          | None                                     | None                                     |
| PSMD14 O00487-1 ENST0000409682       | 26S proteasome non-ATPase regulatory subunit 14 | 26S proteasome subunit. Metalloprotease that specifically cleaves “Lys-63” linked polyubiquitin chains. Plays a role in DSBs and in recombination repair by promoting RAD51 loading. | GO:1905114, cell surface receptor signaling pathway involved in cell-cell signaling | 310            | No            | 65–97                                    | 0.515                          | None                                     | None                                     |
| PURA Q06116-1 ENST000003111           | Poly(U)-binding-splicing factor PURA | DNA and RNA binding, involved in several nuclear processes such as pre-mRNA splicing, apoptosis, and transcription regulation. Binds to poly(U) RNA. | GO:0000398, mRNA splicing via spliceosome | 559            | No            | 90–164                                   | 0.5175                         | 1                                        | 5                                        |
| RAB2A P63019-1 ENST0000262646         | Ras-related protein Rab-2A           | Required for transport from ER to Golgi | GO:0046907, intracellular transport | 212            | No            | 8–46                                     | 0.563                          | None                                     | None                                     |
| RAC1 P63000-1 ENST0000348055          | Ras-related C3 botulinum toxin substrate 1 | GTPase that cycles between GTP active and GDP inactive and plays a role in secretory processes, phagocytosis of apoptotic cells, epithelial cell polarization, neurons adhesion, migration and differentiation, and growth-factor induced formation of membrane ruffles. | GO:1905114, cell surface receptor signaling pathway involved in cell-cell signaling | 192            | No            | 141–171*                                 | 0.236                          | None                                     | 9                                        |
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|--------------------------------------|----------------|----------|----------------------|----------------|--------------|------------------------------------------|-------------------------------------|--------------------------------------------|---------------------------------------------|
| RAN P62838-1 ENST0000543796          | GTP-binding nuclear protein Ran | GTPase involved in nucleocytoplasmic import/export. Required for normal progression through mitosis. | GO:0071426, ribonucleoprotein complex export from nucleus | 216            | No           | 12–50, 118–187                           | 0.179                               | None                                       | None                                        |
| RBBP4 Q90208-1 ENST0000373493        | Histone-binding protein RBBP4  | Core histone binding subunit. Chromatin remodeling. Component of CAF-1, HDAC, NuRD, PRC2, and NURF. | GO:003044, ATP-dependent chromatin remodeling | 425            | No           | 12–63, 228–260, 294–333, 335–384         | 0.3305                              | None                                       | None                                        |
| RBBP5 Q15391-1 ENST0000264585        | Retinoblastoma-binding protein 5 | Plays crucial role in differentiation potential in embryonic stem cells. Gene regulation. Stimulates histone methyltransferases. | GO:0016569, covalent chromatin modification | 538            | No           | 1–18                                      | 0.716                               | None                                       | None                                        |
| RBBP7 Q6576-1 ENST0000380877         | Histone-binding protein RBBP7  | Core histone binding subunit that may target histone remodeling factors. Component of some histone remodeling complexes. | GO:0006338, chromatin remodeling | 425            | No           | 122–156*                                  | 0.5405                              | None                                       | None                                        |
| RBM0 P9817-1 ENST0000037704          | RNA-binding protein 10        | mRNA processing | GO:0050684, regulation of mRNA processing | 930            | No           | 332–375                                   | 0.7125                              | None                                       | 4                                            |
| RBM2 Q9NW64-1 ENST000019981          | Pre-mRNA-splicing factor RBM2 | Required for pre-mRNA splicing as component of the activated spliceosome. | GO:0000398, mRNA splicing, via spliceosome | 420            | No           | 20–50                                     | 0.798                               | None                                       | None                                        |
| RBM3 P9817-1 ENST0000037659          | RNA-binding protein 3         | RNA binding | GO:0050684, regulation of mRNA processing | 157            | No           | 1–20                                      | 0.853                               | None                                       | None                                        |
| RBM3 Q14498-1 ENST0000253363         | RNA-binding protein 39        | Acts as pre-mRNA splicing factor. | GO:0008380, RNA splicing | 530            | No           | 373–403                                   | 0.655                               | None                                       | None                                        |
| RBMX2 Q9Y388-1 ENST0000035536         | RNA-binding motif protein, X-linked 2 | Involved in pre-mRNA splicing as component of spliceosome. | GO:0008380, RNA splicing | 322            | No           | 1–17                                      | 0.907                               | 1                                           |
| RBMY1A1 P0DJT-1 ENST00000382707       | RNA-binding motif protein, Y chromosome, family 1 member A1 | mRNA binding | GO:0050684, regulation of mRNA processing | 496            | No           | 99–130                                    | 0.934                               | None                                       | None                                        |
| RBMY1A1 P0DJT-1 ENST00000382707       | RNA-binding motif protein, Y chromosome, family 1 member A1 | mRNA binding | GO:0050684, regulation of mRNA processing | 496            | No           | 99–130                                    | 0.934                               | None                                       | None                                        |
| RHOA P61586-1 ENST00000418115         | Transforming protein RhoA     | GTPase involved in cytoskeleton organization. Regulates KCNA2. Can be activated by CaMKII. | GO:005114, cell surface receptor signaling pathway involved in cell-cell signaling | 193            | No           | 1–77                                      | 0.268                               | 8                                           | 9                                            |
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|--------------------------------------|-----------------|----------|-----------------------|----------------|--------------|------------------------------------------|---------------------------------|---------------------------------|---------------------------------|
| RHOB P62745-1 ENST00000272233      | RhôB-related GTP-binding protein RHôB      | Mediates apoptosis in neoplastically transformed cells after DNA damage. Myosin contractile ring formation during cell cycle cytokinesis. | GO:00002278: mitotic cell cycle                  | 196            | No          | 18–48                                   | 0.566                           | None                            | None                            |
| RPL10 P27635-1 ENST00000424325     | 60S ribosomal protein L10                  | Component of large ribosomal subunit. May play a role in embryonic brain development. | GO:00021790: embryo development                  | 214            | No          | 48–78                                   | 0.405                           | 1                               | 8                               |
| RPL36A P38381-1 ENST0000053310     | 60S ribosomal protein L36a                | Ribosomal protein | GO:0002181: cytoplasmic translation | 106            | No          | 81–106                    | 0.580                           | None                            | None                            |
| RPS28 P62857-1 ENST00000600659     | 40S ribosomal protein S28                | NA                                   | GO:0006413: translational initiation          | 69             | No          | 54–69                                   | 0.5585                          | None                            | 1                               |
| RPS6KA3 P51812-1 ENST00000379365   | Ribosomal protein S6 kinase alpha-3     | Serine/threonine kinase downstream of ERK. Regulates translation. Modulates mTOR signaling. Role in other pathways. | GO:0004643: protein phosphorylation            | 740            | No          | 114–150, 457–494, 559–595, 681–711   | 0.5825                          | 1                               | Many                            |
| RRAGA Q7L23-1 ENST00000380527      | Ras-related GTP-binding protein A        | Guanine nucleotide binding protein that plays an important role in mTORC1 signaling for amino acid availability. May lead to cell death through TNF-α signaling. | GO:0043200: response to amino acid             | 313            | No          | 16–46                                   | 0.5725                          | None                            | None                            |
| RRM2 P31350-1 ENST00000304367      | Ribonucleoside diphosphate reductase subunit M2 | Provides the precursors necessary for DNA synthesis. | GO:00019438: aromatic compound biosynthetic process | 389            | No          | 345–378                    | 0.769                           | None                            | None                            |
| RTF1 Q9541-1 ENST00000389629       | RNA polymerase-associated protein RTF1 homolog | Component of Pafl complex. Implicated in regulation of development of embryonic stem cell pluripotency. Required for Wnt and Hox genes. | GO:1905114: cell surface receptor signaling pathway involved in cell–cell signaling | 710            | No          | 688–710      | 0.672                           | None                            | None                            |
| RYR2 Q97361-1 ENST00000366374      | Ryanodine receptor 2                     | Mediates calcium release from the sarcoplasmic reticulum and plays a critical role in cardiac muscle contraction. | GO:00030001: metal ion transport             | 4,967          | Yes         | 4,856–4,889                  | 0.841                           | 1                               | Many                            |
| SAT1 P21673-1 ENST00000379270      | Diamine acetyltransferase 1              | Acetylation of small molecule polyamines (e.g., spermidine) | GO:0009058: biosynthetic process           | 171            | No          | 148–171                   | 0.533                           | None                            | None                            |

(Continues)
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|--------------------------------------|----------------|----------|----------------------|----------------|---------------|----------------------------------------|-------------------------------|---------------------------------|---------------------------------|
| SCN2A Q9250-1 ENST0000375437         | Sodium channel protein type 2 subunit alpha (NaV1.2) | Voltage dependent release of sodium permeability. Implicated in hippocampal replay occurring with sharp wave ripples. | GO:003000, metal ion transport | 2,005          | Yes           | 404–414, 854–885                        | 0.7285                        | 1                              | Many                            |
| SCN8A Q9UQ1-1 ENST000035434          | Sodium channel protein type 8 subunit alpha (NaV1.6) | Voltage dependent sodium ion channel | GO:003000, metal ion transport | 1,980          | Yes           | 396–459, 837–875, 910–961, 1,287–1,323, 1,449–1,499, 1,638–1,671, 1,680–1,716, 1,746–1,776 | 0.639                          | Many                           | Many                            |
| SF1 Q1563-1 ENST000037730           | Splicing factor 1 | Required for first step in ATP dependent spliceosome assembly | GO:0000387, spliceosomal snRNP assembly | 639            | No            | 223–258                                 | 0.751                         | None                           | None                            |
| SF3B1 O7553-1 ENST0000333508         | Splicing factor 3B subunit 1 | Pre-mRNA splicing as part of SF3B complex. | GO:0000245, spliceosomal complex assembly | 1,304           | No            | 537–573, 816–848, 962–1,002, 1,005–1,062, 1,133–1,170, 1,189–1,236 | 0.567                          | None                           | 18                              |
| SF3B4 Q1542-1 ENST0000271628         | Splicing factor 3B subunit 4 | mRNA splicing | GO:0005084, regulation of mRNA processing | 424            | No            | 1–23                                    | 0.642                         | 1                              | 4                               |
| SIN3A Q96ST3-1 ENST0000394947        | Paired amphipathic helix protein Sin3a | Transcriptional repressor Regulates cell cycle progression. Required for cortical neuron differentiation and callosal axon elongation. | GO:00099790, embryo development | 1,273           | No            | 112–156                                  | 0.778                         | None                           | 4                               |
| SLC25A5 P0544-1 ENST0000317881       | ADP/ATP translocase 2 | ADP/ATP antiporter that mediates ATP synthesis in the mitochondria. | GO:0005085, transmembrane transport | 298            | Yes           | 283–288                                  | 0.714                         | None                           | None                            |
| SLC9A6 Q9258-1 ENST0000370688        | Sodium/hydrogen exchanger 6 | Exchange of protons for sodium and potassium across endosomes. Contributes to calcium homeostasis. | GO:003000, metal ion transport | 699            | Yes           | 334–371                                 | 0.758                         | None                           | Many                            |
| SMARCA2 F5135-1 ENST000038233        | Probable global transcription activator SNF2L2 | Component of SWI/SNF complex which carries out chromatin remodeling. Also belongs to the neural progenitors-specific chromatin remodeling complex (mBAF complex) and the neuron-specific chromatin remodeling complex (nBAF complex). | GO:0006338, chromatin remodeling | 1,590           | No            | 933–969                                  | 0.634                         | 1                              | Many                            |
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|--------------------------------------|-----------------|----------|----------------------|----------------|---------------|------------------------------------------|-------------------------------------|---------------------------------------------|-----------------------------------------------|
| SMARCA4 P51532-1 ENST00000344626     | Protein         | Involved in chromatin remodeling, part of SWI/SNF complex | GO:0006338, chromatin remodeling | 1,647          | No            | 754–799, 879–912, 955–987, 1,035–1,067   | 0.559                                      | 1                                           | Many                                          |
| SMARCA5 O60264-1 ENST00000283331     | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5 | Helicase that has ATP-dependent nucleosome-remodeling activity. Component of ISWI. Binds to histones | GO:0004344, ATP-dependent chromatin remodeling | 1,052          | No            | 290–321                                  | 0.651                                      | None                                        | None                                          |
| SMARCE1 Q6691-1 ENST00000348513      | SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily E member 1 | Chromatin remodeling to activate or repress genes. Component of SWI/SNF | GO:0006338, chromatin remodeling | 411            | No            | 54–109                                   | 0.785                                      | 5                                           | Many                                          |
| SMCA1 Q4683-1 ENST00000322213        | Structural maintenance of chromosomes protein 1A | Central component of the cohesion complex, which is essential for cohesion of sister chromatids after DNA replication. Involved in DNA repair. | GO:0007059, chromosome segregation | 1,233          | No            | 36–73, 290–321, 636–670, 1,103–1,153    | 0.5205                                     | 4                                           | Many                                          |
| SNA12 O43628-3 ENST00000020945       | Zinc finger protein SNA12 | Transcriptional repressor. Involved in neural development. | GO:0003156, regulation of histone modification | 268            | No            | 202–241                                  | 0.838                                      | None                                        | 2                                             |
| SNRPC P09234-1 ENST00000244520       | U1 small nuclear ribonucleoprotein C | Component of U1 snRNP spliceosome | GO:00008380, RNA splicing | 159            | No            | 9–47                                     | 0.689                                      | None                                        | None                                          |
| SNX12 Q9UMY4-2 ENST00000037474       | Sorting nexin 12| May be involved in intra-cellular trafficking | GO:00051049, regulation of transport | 162            | No            | 6–37                                     | 0.691                                      | None                                        | None                                          |
| SP3 Q02447-1 ENST00000310085         | Transcription factor Sp3 | Translation factor that can act as an activator or repressor depending on isoform or PTM. Binds to GT and GC boxes. Cell cycle regulation, hormone induction, and house-keeping | GO:00009790, embryo development | 781            | No            | 625–658                                  | 0.8405                                     | None                                        | None                                          |
| SPIN1 Q0V657-1 ENST00000375859       | Spindlin-1 | Chromatin reader. Activator of Wnt. May play a role in cell-cycle regulation during transition from gamete to embryo. | GO:1905114, cell surface receptor signaling pathway involved in cell–cell signaling | 262            | No            | 83–115, 240–262                          | 0.616                                      | None                                        | None                                          |
| SPOP O43701-1 ENST00000393328        | Speckle-type POZ protein | Component of Cullin ring based BCR E3 ubiquitin ligase. | GO:0016567, protein ubiquitination | 374            | No            | 19–60, 62–124                            | 0.453                                      | 6                                           | 15                                            |

(Continues)
| Gene symbol, UniProt ID, transcript ID | Encoded protein                     | Function                                                                 | GO pathway or process                                                                 | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|--------------------------------------|------------------------------------|--------------------------------------------------------------------------|----------------------------------------------------------------------------------------|----------------|----------------|-------------------------------------------|---------------------------------------|---------------------------------------------|---------------------------------------------|
| SRF P11831-1, ENST00000265354        | Serum response factor              | Transcription factor that binds to serum response element. Together with MRTFA is coupled to cytoskeletal expression and dynamics. Required for cardiac differentiation and maturation. | GO:00009790, embryo development                                                      | 508            | No             | 134–168                                   | 0.785                                 | None                                        | None                                        |
| SRSF10 P75494-1, ENST00000492112     | Serine/arginine-rich splicing factor 10 | Pre-mRNA splicing                                                        | GO:0050684, regulation of mRNA processing                                                | 262            | No             | 1–44, 32–96                               | 0.79                                  | None                                        | None                                        |
| SRSF2 Q01130-1, ENST00000392485      | Serine/arginine-rich splicing factor 2 | Splicing of pre-mRNA                                                     | GO:0050684, regulation of mRNA processing                                                | 221            | No             | 1–17                                       | 0.603                                 | None                                        | None                                        |
| SRSF3 P4403-1, ENST00000373715       | Serine/arginine-rich splicing factor 3 | RNA binding and pre-mRNA cleavage                                        | GO:0050684, regulation of mRNA processing                                                | 164            | No             | 24–56                                      | 0.390                                 | None                                        | None                                        |
| SRY Q05066-1, ENST00000383070        | Sex-determining region Y protein    | Transcriptional regulator that controls a genetic switch in male development. | GO:0030238, male sex determination                                                      | 204            | No             | 129–163                                   | 0.9095                                | None                                        | 16                                          |
| STAG2 Q8N3U4-1, ENST0000037180       | Cohesin subunit 5A-2                | Component of cohesin complex. Required for cohesion of sister chromatids after DNA replication. | GO:00007059, chromosome segregation                                                     | 1,231          | No             | 113–145                                   | 0.700                                 | None                                        | 5                                           |
| SUMO2 P61996-1, ENST00000420826      | Small ubiquitin-related modifier 2 | Ubiquitin like protein that can be attached to proteins on lysine residues. | GO:00018205, poly(ADP-ribosyl) lysine modification                                      | 95             | No             | 17–54                                     | 0.231                                 | None                                        | None                                        |
| SUZ12 Q15022-1, ENST00000322282      | Polycomb protein SUZ12             | Polycomb group protein. Involved in histone methylation.                  | GO:0034968, histone lysine methylation                                                   | 739            | No             | 308–342                                   | 0.811                                 | None                                        | 2                                           |
| TAF1 P21675-1, ENST00000423759       | Transcription initiation factor TFIIID subunit 1 | Largest component and core scaffold of the TFIIID basal transcription factor complex. Kinase and histone acetyltransferase activity. | GO:00016573, histone acetylation                                                        | 1,872          | No             | 1,328–1,370                               | 0.738                                 | 2                                          | 25                                          |
| TAOK1 Q7KX7-1, ENST00000261786       | Serine/threonine-protein kinase TA01 | Serine/threonine-protein kinase involved in MAPK cascade, DNA damage response, and regulation of cytoskeleton stability | GO:00070507, regulation of microtubule cytoskeleton organization                         | 1,001          | No             | 181–215                                   | 0.742                                 | None                                        | None                                        |
| TBC1D3H Q0C7X7-1, ENST00000455054    | TBC1 domain family member 3H       | Acts as a GTPase activating protein for RAB5.                              | GO:0006886, intracellular protein transport                                              | 549            | No             | 344–376 (based on ENST00000455054)         | 0.986                                 | None                                        | None                                        |
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|--------------------------------------|----------------|----------|-----------------------|----------------|---------------|-------------------------------------------|-------------------------------------|-----------------------------------|------------------------------------------|
| TBL1XR1 Q9BZK7-1 ENST00000430069     | F-box-like/WD repeat-containing protein TBL1XR1 | Recruitment of ubiquitin/19S proteasome to nuclear receptor-regulated transcription units. Probably acts as integral component of the N-Cor corepressor complex. | GO:0009790, embryo development | 514 | No | 261–284, 322–355 | 0.588 | None | Many |
| TCF4 P15844-1 ENST00000564999        | Transcription factor 4 | Transcription factor that binds to immunoglobulin enhancer. Involved in neuron differentiation. | GO:0006366, transcription by RNA polymerase II | 667 | No | 552–587 | 0.732 | 1 | Many |
| THOC2 Q8N1Z7-1 ENST00000245838       | THO complex subunit 2 | Required for efficient export of poly-A spliced mRNA. Component of TRX complex. | GO:0006397, mRNA processing | 1,593 | No | 135–188, 698–731, 733–786, 1,089–1,100 | 0.653 | 1 | 17 |
| TLK2 Q86UE8-1 ENST00000326270        | Serine/threonine-protein kinase tousled-like 2 | Serine/threonine kinase involved in chromatin assembly | GO:1902275, regulation of chromatin organization | 772 | No | 651–689 | 0.631 | None | 2 |
| TOPI P11387-1 ENST00000361337        | DNA topoisomerase 1 | Topoisomerase. DNA repair/strain resolution. | GO:0009790, embryo development | 765 | No | 476–512 | 0.700 | None | 2 |
| TRA2B P6295-1 ENST00000453386        | Transformer-2 protein homolog beta | mRNA splicing | GO:050684, regulation of mRNA processing | 288 | No | 104–138 | 0.757 | None | None |
| TRIM24 OL5164-1 ENST00000345326      | Transcription intermediary factor 1-alpha | Transcriptional coactivator that interacts with numerous nuclear receptors and modulates transcription. Interacts with chromatin histone H3 modifications. | GO:0006351, transcription, DNA-templated | 1,050 | No | 818–858 | 0.854 | None | None |
| TRPC5 Q5UL62-1 ENST0000026239        | Short transient receptor potential channel 5 | Calcium channel. Causes neuron apoptosis. | GO:00030001, metal ion transport | 973 | Yes | 290–323 | 0.716 | None | None |
| TUBA1A Q71134-1 ENST00000301071       | Tubulin alpha-1A chain | Tubulin chain | GO:0000226, microtubule cytoskeleton organization | 451 | No | 1–41, 41–172, 174–241, 243–256, 258–338, 340–399, 401–439 | 0.000 | Many | Many |
| TUBA1B P64863-1 ENST00000336023       | Tubulin alpha-1B chain | Tubulin chain | GO:00007017, microtubule-based process | 451 | No | 1–18, 82–117, 139–160, 185–233, 243–266, 341–374, 382–441 | 0.16 | None | None |
| TUBB P57417-1 ENST00000327892         | Tubulin beta chain | Major component of microtubules | GO:0000226, microtubule cytoskeleton organization | 444 | No | 49–120, 122–157, 190–240, 242–273, 297–330 | 0.171 | 5 | 14 |

(Continues)
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|----------------------------------------|-----------------|----------|----------------------|----------------|--------------|-----------------------------------------|-------------------------------------|------------------------------------------|---------------------------------------------|
| TUBB2A Q13885-1 ENST00000333628        | Tubulin beta-2A chain | Tubulin component | GO:0000226, microtubule cytoskeleton organization | 445 | No | 300–367 | 0.256 | 1 | 17 |
| TUBB2B Q9HVA1-1 ENST0000259818         | Tubulin beta-2B chain | Tubulin component Implicated in neuronal migration | GO:00009790, embryo development | 445 | No | 283–356 | 0.256 | 2 | Many |
| TUBB4B P68371-1 ENST0000034034         | Tubulin beta-4B chain | Subunit of microtubules | GO:00007077, microtubule-based process | 445 | No | 163–198 | 0.257 | None | None |
| U2AF1 Q1081-1 ENST00000291352          | Splicing factor U2AF35 kDa subunit | Plays critical role in mRNA splicing. | GO:00008380, RNA splicing | 240 | No | 1–32 | 0.446 | 2 | 4 |
| U2AF2 P68381-1 ENST0000030924          | Splicing factor U2AF65 kDa subunit | Pre-mRNA splicing | GO:00008380, regulation of mRNA processing | 475 | No | 188–220, 245–309 | 0.4625 | None | None |
| U2SURP Q130421 ENST0000047385          | U2 snRNP-associated SURP motif-containing protein | RNA binding | GO:00008380, RNA splicing | 1,029 | No | 291–325, 585–621 | 0.7705 | None | None |
| UBC P199246-1 ENST00000536769          | Polyubiquitin-C | Ubiquiti | GO:1905114, cell surface receptor signaling pathway involved in cell–cell signaling | 685 | No | 470–509 | 0.455 | None | None |
| UBE2D3 P61077-1 ENST00000453744        | Ubiquitin-conjugating enzyme E2 D3 | E2 ubiquitin enzyme | GO:00006513, protein monoubiquitination | 147 | No | 33–98, 130–147 | 0.178 | None | None |
| UBE2E3 Q96794-1 ENST00000410062        | Ubiquitin-conjugating enzyme E2 E3 | Accepts ubiquitin from E1 complex. Participates in regulation of transepithelial sodium transport in renal cells | GO:00016567, protein ubiquitination | 207 | No | 65–95 | 0.546 | None | None |
| UBE2H P62256-1 ENST00000355621          | Ubiquitin-conjugating enzyme E2 H | E2 ubiquitin ligase | GO:00000209, protein polyubiquitination | 183 | No | 33–67 | 0.326 | None | 1 |
| UBE2I P63279-1 ENST00000355803          | SUMO-conjugating enzyme UBC9 | Covalently attaches SUMO to target proteins. | GO:00018205, poly(ADP-ribosyl)-lysine modification | 158 | No | 62–98, 117–148 | 0.193 | None | None |
| UBE2K P61086-1 ENST00000261427          | Ubiquitin-conjugating enzyme E2 K | E2 ubiquitin ligase | GO:00000209, protein polyubiquitination | 200 | No | 1–23 | 0.370 | None | None |
| UHRF2 Q96PU4-1 ENST00000276893          | E3 ubiquitin-protein ligase UHRF2 | E3 ubiquitin ligase that plays important roles in DNA methylation, histone modifications, cell cycle, and | GO:00016567, protein ubiquitination | 802 | No | 582–626 | 0.721 | None | None |
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|----------------------------------------|-----------------|----------|---------------------------|----------------|----------------|------------------------------------------|----------------------------------------|---------------------------------------------|---------------------------------------------|
| USP9Y O00507-1 ENST0000338981          | Probable ubiquitin carboxyl-terminal hydrolase FAF-Y | Probable deubiquitinase | Essential component of TGF-
Beta/BMP signaling. | 2,555 | No | 1,326–1,381 | 0.951 | None | None |
| UTY O14607-1 ENST0000331397            | Histone demethylase UTY | Male specific histone demethylase | GO:0016570; histone modification | 1,347 | No | 124–158, 873–918, 1,094–1,138, 1,194–1,225 | 0.841 | None | None |
| VAV1 P1548-1 ENST0000602342           | Proto-oncogene vav | Couples to tyrosine kinase signals with rho/Rac GTPases, and leads to cell differentiation and/or proliferation. | GO:0010942; Positive regulation of cell death | 845 | No | 365–401 | 0.778 | None | 3 |
| WNK3 Q9BYP7-1 ENST0000354646          | Serine/threonine-protein kinase WNK3 | Serine/threonine kinase that plays an important role in electrolyte homeostasis. | GO:0043270; positive regulation of ion transport | 1,800 | No | 332–364 | 0.915 | None | 2 |
| XPR1 Q9UBH6-1 ENST0000367800          | Xenotropic and polyclonotypic retrovirus receptor 1 | Phosphate homeostasis. Phosphate export. Binds inositol polyphosphates. | GO:0006873; cellular ion homeostasis | 696 | Yes | 112–149 | 0.790 | 1 | 5 |
| YTHDC1 Q66MU7-1 ENST0000344157        | YTH domain-containing protein 1 | Pre-mRNA splicing; mRNA export; involved in spermatogenesis. | GO:0050684, regulation of mRNA processing | 727 | No | 361–395 | 0.872 | None | None |
| YY1 P2548-1 ENST0000262238            | Transcriptional repressor protein YY1 | Transcription factor that exhibits positive and negative control on large number of genes. Binds to CCGCCA/INTT. | GO:0001358; regulation of cell growth | 414 | No | 288–326, 338–440 | 0.526 | 4 | 5 |
| ZBTB16 Q05516-1 ENST0000335953        | Zinc finger and BTB domain-containing protein 16 | Transcriptional repressor. May play a role in myeloid maturation. | GO:00099790, embryo development | 673 | No | 576–622 | 0.7855 | None | 1 |
| ZBTB19 Q9HC78-1 ENST0000474710        | Zinc finger and BTB domain-containing protein 20 | May be a transcription factor involved in hematopoiesis, oncogenesis and, immune response. | GO:0001678; cellular glucose homeostasis | 741 | No | 574–618 | 0.7355 | 9 | 19 |
| ZEB2 O60315-1 ENST0000558370          | Zinc finger E-box-binding homeobox 2 | Transcriptional inhibitor of E-cadherin and represses expression of MEOX2. Binds to CACCT in different promoters. | GO:00099790, embryo development | 1,214 | No | 296–326, 1,064–1,094 | 0.786 | 1 | Many |
| ZFX P1700-1 ENST0000339777            | Zinc finger X-chromosomal protein | Probably a transcriptional activator. | GO:0006357; regulation of transcription by RNA polymerase II | 805 | No | 414–444 | 0.720 | None | None |

(Continues)
| Gene symbol, UniProt ID, transcript ID | Encoded protein | Function | GO pathway or process | Protein length | Transmembrane? | Intolerant segment(s) (UniProt numbering) | Median MTR score for entire protein | No. of ClinVar variants in intolerant segment(s) | No. of ClinVar variants in the whole protein |
|---------------------------------------|----------------|----------|----------------------|---------------|----------------|------------------------------------------|---------------------------------|---------------------------------|---------------------------------|
| ZFY P08048-1 ENST00000383052          | Zinc finger Y-chromosomal protein | Probable transcription factor | GO:0006357, regulation of transcription by RNA polymerase II | 801 | No | 654–691c | 0.828 | None | None |
| ZMAT2 Q6NCO1-1 ENST00000274712        | Zinc finger matrin-type protein 2 | Involved in pre-mRNA splicing as a component of the spliceosome. | GO:0000398, mRNA splicing, via spliceosome | 199 | No | 67–108 | 0.709 | None | None |
| ZMYM3 Q4202-1 ENST00000314425         | Zinc finger MYM-type protein 3 | Plays a role in cell morphology and cytoskeletal organization. | GO:0007010, cytoskeleton organization | 1,370 | No | 1,185–1,215 | 0.724 | None | 3 |
| ZMYND8 Q9ULU41-1 ENST00000311275      | Protein kinase C-binding protein 1 | Transcriptional co-repressor for KDM5D. Function seems to be histone recognition. | GO:0006284, regulation of cell development | 1,186 | No | 1,033–1,072b | 0.790 | None | None |
| ZNF84 P51523-1 ENST00000327689        | Zinc finger protein 84 | May be involved in transcription | GO:0006357, regulation of transcription by RNA polymerase II | 738 | No | 486–527 | 0.885 | None | None |

The transcript containing the intolerant segment is not found in UniProt and is not the canonical transcript. The amino acid numbering provided for the zero-tolerance segment is based on the sequence of the protein encoded by the indicated transcript.

The transcript containing the intolerant segment is not found in UniProt. The amino acid numbering provided for the zero-tolerance segment is based on the sequence of the protein encoded by the indicated transcript.

UniProt numbering, which in this case is different from the numbering of the MTR-designated canonical transcript in gnomAD.

UniProt transcript used that is not considered the canonical sequence by UniProt.

Abbreviations: ATP, adenosine triphosphate; ER, endoplasmic reticulum; GO, gene ontology; GPCR, G-protein coupled receptor; GDP, guanosine diphosphate; GTP, guanosine triphosphate; MTR, missense tolerance ratio; PTM, post-translational modification.
median MTR score is presented for all segments within each protein.

In addition to the 257 proteins with certain zero-tolerance segments, we also found 33 human proteins that have 31 or longer residue segments with an MTR score of 0, but for which the statistics associated with this score are uncertain because of an insufficient number of observed silent mutations with the intolerant segments. These 33 proteins are listed in Table S1 and will require additional data to determine whether the preliminary MTR = 0 score seen for at least one segment within each of these proteins is confirmed in a statistically robust manner. It may be significant that 7 of these 33 proteins have sites of known ClinVar variants, suggestive of high relevance to human health (Table S1).

2.2 Homology of human zero-tolerance proteins to corresponding proteins from other mammals

For each of the 257 proteins containing one or more zero-tolerance segments, we conducted BLASTP sequence homology searches for both the entire protein sequence and the zero-tolerance segments. Results were analyzed for the 250 closest mammalian homologs. Figure 2 gives a representative sample of the results, while Figure S1 shows the results for all 257 proteins. For each protein the statistical distribution of sequence identity to the closest 250 mammalian homologs is presented for both full-length protein sequence (black) and zero-tolerance segment(s) (red).

For most proteins, it is seen that the median of the distribution of % sequence homology is much higher for the zero-tolerance segments than it is for the entire protein sequence. Indeed, for a great many zero-tolerance segments, the degree of sequence identity to all 250 closest mammalian homologs is 100%. However, there are also a number of exceptions seen in Figures 2 and S1, where the median sequence homology observed for a given zero-tolerance segment is significantly less than 100%. For example, the intolerant segment found in CTCF exhibits a lower median homology score to mammalian homologs than does the full protein sequence of that protein.

There are a variety of potential explanations for why any given zero-tolerance segment is not absolutely conserved among mammalian homologs. Because currently there are only roughly 150 fully sequenced mammalian genomes, the fact that we are considering the data for the 250 nearest homologs implies that many of the homologs included in the analysis for a given protein are paralogs, not orthologs. Between paralogs, even functionally critical residues are sometimes expected to exhibit variation. Another and particularly intriguing possibility is that some less-than-100%-conserved human zero-tolerance segments play critical roles in establishing traits that are unique to humans. Only careful future studies of specific instances will provide convincing explanations for why some of the proteins shown in Figures 2 and S1 contain human zero-tolerance segments that are less-than-100% conserved.

2.3 Likely protein basis for most evolutionary intolerance associated with protein-encoding genes

While it cannot be ruled out for all entries in Table 1 that the mechanism responsible for evolutionary purifying selection involves changes in parent DNA or mRNA structure (see the previous review), we hypothesize that for the vast majority of cases, evolutionary intolerance stems from the altered properties of the encoded mutant protein. This is here supported by two observations. (a) A number of the proteins listed in Table 1 are known to directly form complexes with other proteins appearing in
Figure 2. Representative examples of sequence identity patterns for proteins containing zero-tolerance segments, comparing both the whole-protein (black plots) and the intolerant segmental (red plots) homology levels to the 250 nearest mammalian homologs following BLASTP searches of NCBI. GENE.1, GENE.2, and so forth indicate which non-contiguous intolerant segment for that gene was searched. The distributions of sequence identities seen for the 250 closest homologs to each protein are presented as box-and-whiskers plots. The bold bar is the median, the wings of the bars are the quartiles and the whiskers are 1.5 times the inner quartile ranges. The dots are outliers that lie beyond the whiskers. The complete results for all 257 proteins with zero-tolerance segments are presented in Figure S1.
this table, suggesting that disruption by a single mutation in a single subunit of critical multi-protein complexes is a common mechanism of underlying zero-tolerance. In some other cases, multiple proteins containing zero-tolerance segments are seen to be on the same pathways, even if they do not actually form a complex. (b) There are many proteins appearing in Table 1 that are known to be central players in human biology and physiology—proteins that one might expect to contain intolerant segments. These include calmodulin, ubiquitin, SUMO, clathrin, various tubulin subunits, actin, and the ryanodine receptor. In light of these considerations, this paper focuses on the implications of genetic intolerance as it relates to the encoded proteins.

2.4 | A case study of intolerance: The voltage-gated potassium and sodium channels

Although a comprehensive structural study of all 257 proteins with zero-tolerance regions is beyond the scope of the current work, we nonetheless subjectively perused several dozen of the proteins in Table 1. Results suggest that zero-tolerance segments tend to occupy well-structured regions of proteins, often including functionally-critical sites. An example is provided by the six voltage-gated potassium channels and two voltage-gated sodium channels appearing in Table 1 (see list in Table S2). With only one exception, the 19 intolerant segments documented for these eight channels are contained within the part of the channels that spans the critical transmembrane S4 segment of the voltage sensor domain through the transmembrane S6 segment, the latter of which includes the channel gate and ends the pore domain (see Table S2). The structural elements in this span are all known to be critical to voltage-gated sodium and potassium channel function, where S4 and the S4–S5 linker are central to channel regulation by the transmembrane electrical potential. The actual pore is comprised of S5, the selectivity filter, the pore helix, and S6.16,17 The location of zero-tolerant segments in these structural elements strongly implicates mutation-induced alteration of channel function as the mechanistic basis for evolutionary intolerance associated with these segments. It is interesting that from channel to channel the exact location of the intolerant segments varies. For example, the single zero-tolerance segment in KCNA3 spans the S4 segment and S4–S5 linker, which are key for voltage regulation, while the single zero-tolerance segment in KCNH7 spans the pore helix, selectivity filter, and S6, which are critical for ion selectivity and flux.16,17

The single zero-tolerance segment that was not located within the functionally-central S4 through S6 part of the channels is the 82–113 residues segment found in the KCNB1 potassium channel. This segment is located in its N-terminal tetramerization (T) domain, a domain found in some, but not all voltage-gated potassium channels. Mutagenesis studies of H105 located within this intolerant segment revealed that mutations at this site do not interfere with KCNB1 homotetramerization, but rather disrupt heterotetramerization with subunits of voltage-gated K,6 potassium channel family members.18 This strongly suggests that the basis for zero-tolerance in this segment is not disruption of the formation of homotetrameric KCNB1 channels but rather disruption of the formation of heterotetrameric KCNB1/K,6 channels.

A final observation should be made about the sodium channel SCN2A. Unlike homotetrameric potassium channels, human voltage-gated sodium channels combine all four subunits in a single long chain in which the four connected “pseudo-subunits” are homologous, but are not identical in sequence, resulting in a fourfold semi-symmetric channel.19 It is interesting that only two of the pseudo-subunits of SCN2A contain zero-tolerance segments, not all four. Some pseudo-subunits in voltage-dependent sodium channels are evidently more tolerant of mutations than others.

2.5 | Previously overlooked proteins containing zero-tolerance segments

While the zero-tolerance proteins include a number of prominent proteins, on the opposite end of the spectrum are a number of genes/proteins listed in Table 1 that are almost completely uncharacterized. A February 2022 PubMed search on each of the following nine genes yielded, at most, only eight papers mentioning each: CLASRP, GOLGA8G, HMGN4, OR4F17, RBMX2, TBC1D3H, U2SURP, ZMAT2, and ZNF84. The presence of zero-tolerance segments within these proteins suggests that at least some of them are associated with critical physiological functions and/or pathophysiology. While only further study will confirm this prediction, the MTR data seems compelling that such studies are merited. Here, we further highlight the case of OR4F17, which is a membrane protein and putative olfactory receptor. Only 36 of the zero-tolerance proteins (13%) are integral membrane proteins, which include the aforementioned voltage-gated channels (Table 1). This is despite the fact that membrane proteins represent roughly 20–30% of all human proteins20 and are the targets for more than 50% of all approved drugs.21 This highlights the fact that the factors that decide what represents a good target for drug
development correlate only partially with the priorities of natural selection. Indeed, a particularly intriguing observation is that while the human G-protein coupled receptor (GPCR) superfamily includes the targets for about one third of all approved drugs, OR4F17 is the only GPCR among the 257 proteins of Table 1 and is classified as one of the 500 human olfactory receptors. This raises the question that why an olfactory receptor would contain a zero-tolerance segment. We suggest three competing hypotheses. First, it could be that mutations in the intolerant segment of this receptor (located at its N-terminus) could result in a toxic gain-of-function effect such as promoting the formation of aggregates or amyloids by this protein. Another possibility is that OR4F17 is not actually an olfactory receptor but has a different and very important physiological function that is disrupted by mutations in its intolerant segment. A third possibility is that it is an olfactory receptor but has additional physiological functions. This would not be unprecedented. Only future experiments will determine which, if any, of these hypotheses are correct. However, this serves as another illustration of the power of intolerance analysis to direct attention to interesting biological questions.

2.6 | Proteins involved in RNA splicing represent the largest group of proteins containing at least one intolerant segment

We sought preliminary insight into which pathways, networks, and protein complexes are most commonly represented among the 257 proteins with intolerant segments. Cytoscape stringApp with a confidence cutoff of 0.95 was used to determine high confidence interactors. This approach yielded protein interaction maps that group proteins based on broad molecular or cellular functional categories (Figure 3). The largest clusters of networked proteins are associated with central cellular processes such as chromatin remodeling, protein degradation, RNA splicing, the cytoskeleton, the cell cycle, and nucleic acids biochemistry.

Further analysis of the protein interaction-mapping presented in Figure 3 using the stringApp network clustering with a granularity parameter of three was used to help identify sets of proteins that may participate in functional complexes (Figure 4). Major complexes include proteins of the spliceosome, translation, tubulin, and NMDA receptor.

![Figure 3](image-url)
GO Panther's overrepresentation analysis was employed to identify biological processes that are over-represented in the intolerant gene list as compared to the *Homo sapiens* reference. We filtered for processes that had less than 500 proteins in the reference list to avoid very general biological processes and instead focused on more specific pathways. In addition, the $p$ value cut-off was required to be less than $5 \times 10^{-10}$. As can be seen in Figure 5a, the most overrepresented biological processes are all or in some way related to mRNA processing, particularly RNA splicing. The two other pathways noted were histone modification and regulation of membrane potential. For further confirmation, the list of proteins was analyzed using Enrichr. Consistent with Panther, it was seen that three different databases (Bioplanet, WikiPathway 2021 Human, and KEGG 2021 Human) indicate that mRNA processing was the most significantly enriched pathway. Approximately 50 of the 290 proteins with zero-tolerance segments were identified by Panther having the gene ontology (GO) term mRNA processing (see Table 1). When GO Panther's overrepresentation analysis results were filtered to retain results only for the proteins with a zero-tolerance segment that is at least 41 residues long, the only two categories yielding $p < 5 \times 10^{-10}$ were RNA splicing and mRNA processing (Figure 5b), further highlighting the robust enrichment of these protein functional categories in Table 1.

The gene enrichment analyses all point to RNA splicing as the biological process that is associated with the largest number of proteins that contain an intolerant segment. This may well reflect the importance of mRNA splicing in early human development (conception to birth), where this process enables proteins to be remodeled to suit varying roles during the developmental phases of human gestation. Failure to express the correct protein isoforms at the right time may be a particularly common mechanism of purifying selective pressure on the responsible gene variations.

### 2.7 | Association of ClinVar variants with proteins having intolerant segments

For each protein with an MTR = 0 segment, we also examined whether there are ClinVar missense variants...
that encode amino acid changes in that protein and recorded this observation in Table 1. As of February 2022, we found that 127 of the 257 proteins contain at least one ClinVar missense mutation encoding an amino acid change in the protein. It is interesting that the other 130 of these proteins have no known or suspected disease mutations associated with them, highlighting the ability of intolerance analysis to detect proteins that evidently may be essential to human reproduction or gestational development but are not associated with known human genetic disorders. Currently, detection of disease variants is usually based on genetic sampling and analysis of people after they have been born, explaining why mutations in such essential genes may have escaped detection.

3 | CONCLUSION

This paper reports that 257 human proteins contain zero-tolerance segments, as identified by MTR analysis. Some of these proteins were previously known to be associated with genetic disorders and some were not. While not all proteins containing zero-tolerance segments can be functionally grouped with other such proteins, about half were found in one of a half dozen functionally-related groups of protein, the largest of which (containing nearly 20% of all zero-tolerance proteins) is associated with RNA splicing and related RNA biochemistry.

We hope that this report of 257 human proteins that contain zero-tolerance segments will motivate studies of these proteins to establish exactly how and why mutations in intolerant segments within each protein result in populations, often sick children with de novo (non-inherited) mutations. These very rare variants may cause or contribute to human disorders, but are not absolutely filtered out of the human population because they do not prevent birth.
purifying selection in the human population. This will require insight into the human physiological role(s) of each protein and also structural and structure–function data (see Perszyk et al. for a recent method that may support such efforts). For some of these proteins, such as the voltage-gated potassium and sodium channels, there may already be enough information in the literature to rationalize the presence of zero-tolerant segments. However, even for these channels, questions remain. For example, mutations in the zero-tolerance segments of the sodium channels SCN2A and SCN8A are subjected to purifying selection even though these mutations would occur under heterozygous WT/mutant expression conditions and even though sodium channels, unlike potassium channels, form monomeric channels. Does this mean a 50% reduction in the function of SCN2A or SCN8A is sufficient to prevent human reproduction or terminate life before birth or is it instead the case that mutations in zero-tolerant segments in these proteins induce some sort of toxic gain-of-function effect that compounds the impact of partial loss-of-function under WT/mutant heterozygous conditions? Future studies may be required to address such questions.

For proteins that have previously escaped significant notice, such as the putative olfactory receptor, OR4F17, observation of a zero tolerant segment suggests a critical and previously overlooked role for these proteins in human reproduction and/or health. Observation of zero-tolerance segments in proteins may be particularly useful as a way of pointing investigators to proteins that are critical for human reproduction and/or pre-birth development, but for which associated causative mutations have never been detected.

Finally, there are other interesting questions triggered by this work. These include the aforementioned question of why some zero-tolerance segments in human protein are not 100% conserved among their nearest mammalian relatives. Another question is inspired by Figure 1a, where it is seen that there is a modest population of proteins that have not only a zero-tolerance segment, but also contain segments with MTR values higher than 1.0, suggesting these latter segments are experiencing evolutionary pressure to rapidly mutate. Does this suggest that such proteins are critical to human reproduction and/or health, yet also are being pressured either to adapt to changes in the human environment, to further optimize a current function, or to acquire a new function or mode of regulation? We hope that addressing questions such as these will ultimately advance our understanding of the molecular biology of human health, reproduction, development, and disease.

4 | MATERIALS AND METHODS

4.1 | MTR analysis

An Excel file containing a well-annotated list of all canonical human genes was provided by Prof. Anthony Capra of the University of California, San Francisco. From this list, we deleted all genes that encode various forms of non-coding RNA, leaving a list of roughly 20,000 protein-encoding genes. Each gene was then subjected to MTR analysis using the web-mounted MTR-Viewer server (http://biosig.unimelb.edu.au/mtr-viewer/). Version 2 of MTR analysis was run using the default window size of 31 residues. This program conducts MTR analysis in “sliding sequence” fashion for each possible 93 nucleotide segment in the coding gene transcript and returns a plot of the segmental MTR score versus the position of the amino acid in the middle of the encoded 31 residues segment. When a residue is within 16 residues of the protein’s N- or C-terminus, analysis is conducted, but in a truncated manner. For example, for residue 10 in any given protein, the reported MTR score will be for the gene segment that encodes residues 1–25. The MTR plots generated by the server for each protein were then manually inspected and the minimum MTR score observed for the analyzed gene/protein was recorded along with the corresponding residue number at the center of the analyzed segment. For proteins having multiple overlapping and/or non-overlapping MTR = 0 segments, the locations of all such segments were recorded. MTR plots revealed that there were 257 human proteins that exhibited at least one statistically robust MTR = 0 segment.

The 257 genes and their encoded proteins that exhibited at least one statistically robust MTR = 0 segment are tabulated (Table 1) with both gene codes and UniProt identifiers (https://www.UniProt.org/). It was found that the canonical transcript for a given gene analyzed by the MTR version 2 program does not always correspond to the canonical protein sequence listed in UniProt, usually because the MTR-analyzed transcript is a splice variant of the transcript that encodes the UniProt-canonical protein. For such instances this is noted in Table 1 and, to avoid confusion, the reported amino acid sequence of the intolerant segment is provided using the residue numbering for the canonical UniProt sequence. There were a few cases where the intolerant segment was not found in the sequence of the UniProt-listed splice variant form(s) of the protein. In these cases, a note is added to the table.

In conjunction with the primary MTR plot for each gene/protein, the output of the MTR-Viewer server also includes a plot of the positions of any known ClinVar33 variants for the analyzed gene/protein. Along with
tabulated MTR data for each protein entry we also included the total number of ClinVar variants in the protein and the number that are located within the MTR = 0 segment(s), if any.

Proteins with intolerant segments were also manually characterized based on their function. The GO terms were tabulated and pathway analysis was also conducted, as described in the following sections. We also recorded whether each protein entry contains a transmembrane domain.

In addition to the 257 proteins with statically robust zero-tolerance segments there were additional 33 proteins that contained MTR = 0 segments, but for which there were not enough observed silent mutations within these segments in gnomAD to ensure that MTR = 0 is statistically robust. These proteins are listed in Table S1, along with additional information regarding the location of the candidate intolerant segment(s) in each protein’s sequence. These 33 proteins remain candidates as having zero-tolerance segments, but more human sequences will be required to increase the number silent mutations to the point where statistically reliable MTR scores can be calculated.

4.2 Sequence homology searches

For each of the 257 protein sequences of Table 1 that contain one or more zero-tolerance segments we ran BLASTP using the default search parameters against all available mammalian protein sequences. For each search we saved the output for the 250 closest mammalian homologs. We also ran BLASTP for each protein’s zero-tolerance segment(s). For each protein, the median % sequence identity for both the full-length sequence and the zero-tolerance segment(s) was determined along with related statistics (Figure 2 and S1).

BLASTP searches were also conducted for the 33 proteins of Table S1 that contain an MTR = 0 segment, but for which the result is not statistically definitive, as summarized in Figure S2.

4.3 Protein–protein interaction analysis

Protein networks for proteins with MTR = 0 segments were constructed using the Cytoscape stringApp with a confidence cutoff of ≥ 0.95 stringdb score. Next, a granulation value of 3 was applied to determine refined complexes. The thickness of the lines connecting protein pairs in the granulated Cytoscape networks was set based on the stringdb experiment scores.

4.4 Pathway analysis

Gene symbols for proteins with at least one MTR = 0 segment were input into GO Panther’s statistical overrepresentation test to determine which biological pathways are overrepresented compared to the reference human gene list (http://www.pantherdb.org/). The processes considered for evaluation were those that encompass less than 500 genes in the human genome reference to filter for GO biological processes functions that are more specific and reduce non-specific overarching GO terms. In addition to this criterion, the GO term must also have $p < 5.0 \times 10^{-10}$. Additionally, the gene list was input into Enrichr (https://maayanlab.cloud/Enrichr/) to determine the biological pathways involved.

AUTHOR CONTRIBUTIONS

Adam Sanders: Data curation (equal); formal analysis (equal); investigation (equal); writing – review and editing (equal). Jake Hermanson: Data curation (equal); formal analysis (equal); investigation (equal); visualization (equal); writing – review and editing (equal). David Samuels: Investigation (equal); writing – review and editing (equal). Lars Plate: Investigation (equal); methodology (equal); supervision (equal); writing – review and editing (equal). Charles Sanders: Conceptualization (equal); funding acquisition (equal); project administration (equal); supervision (equal); writing – original draft (lead); writing – review and editing (lead).

ACKNOWLEDGMENTS

This work was supported by NIH grants R01 NS095989, R01 HL122010, and RF1 AG056147 to Charles R. Sanders and R35 GM133552 to Lars Plate. We thank Anthony Capra of the University of California—San Francisco for kindly providing us with an Excel file tabulating all known human genes.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Lars Plate https://orcid.org/0000-0003-4363-6116
Charles R. Sanders https://orcid.org/0000-0003-2046-2862

REFERENCES

1. Silk M, Petrovski S, Ascher DB. MTR-Viewer: Identifying regions within genes under purifying selection. Nucleic Acids Res. 2019;47:W121–W126.
2. Traynelis S, Miller K, Wang Q, et al. Optimizing genomic medicine in epilepsy through a gene-customized approach to missense variant interpretation. Genome Res. 2017;27:1715–1729.

3. Perszyk RE, Kristensen AS, Lyuboslavsky P, Traynelis SF. Three-dimensional missense tolerance ratio analysis. Genome Res. 2021;31:1447–1461.

4. Fadista J, Oskolkov N, Hansson O, Groop L. LoFtool: A gene intolerance score based on loss-of-function variants in 60,706 individuals. Bioinformatics. 2017;33:471–474.

5. Havrilla JM, Pedersen BS, Layer RM, Quinlan AR. A map of constrained coding regions in the human genome. Nat Genet. 2019;51:88–95.

6. Hayeck TJ, Stong N, Wolock CJ, et al. Improved pathogenic variant localization via a hierarchical model of sub-regional intolerance. Am J Hum Genet. 2019;104:299–309.

7. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–443.

8. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–291.

9. Li R, Roden DM, Capra JA. The 3D mutational constraint on amino acid sites in the human proteome. Nat Commun. 2022;13:3273.

10. Samocha KE, Robinson EB, Sanders SJ, et al. A framework for the interpretation of de novo mutation in human disease. Nat Genet. 2014;46:944–950.

11. Li GC, Forster-Benson ETC, Sanders CR. Genetic intolerance analysis as a tool for protein science. Biochim Biophys Acta Biomembr. 2020;1862:183058.

12. Portelli S, Barr L, de Sa AGC, Pires DEV, Ascher DB. Distinctive regions and potential rescue pharmacology. PLoS Genet. 2021;17:e1006536.

13. Pluznick JL, Protzko RJ, Gevorgyan H, et al. Olfactory receptor responding to gut microbiota-derived signals plays a role in renin secretion and blood pressure regulation. Proc Natl Acad Sci USA. 2013;110:4410–4415.

14. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: Network analysis and visualization of proteomics data. J Proteome Res. 2019;18:623–632.

15. MI H, Ebert D, Muruganujan A, et al. PANTHER version 16: A revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res. 2021;49:D394–D403.

16. Xie Z, Bailey A, Kuleshov MV, et al. Gene set knowledge discovery with Enrichr. Curr Protoc. 2021;1:e90.

17. Huang R, Grishagin I, Wang Y, et al. The NCATS BioPlanet—An integrated platform for exploring the universe of cellular signaling pathways for toxicology, systems biology, and chemical genomics. Front Pharmacol. 2019;10:445.

18. Palshikar MG, Hilchey SP, Zand MS, Thakar J. WikiNetworks: Translating manually created biological pathways for topological analysis. Bioinformatics. 2021;38:869–871.

19. Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. Nat Rev Mol Cell Biol. 2017;18:437–451.

20. Gao C, Wang Y. mRNA metabolism in cardiac development and disease: Life after transcription. Physiol Rev. 2020;100:673–694.

21. Giudice J, Cooper TA. RNA-binding proteins in heart development and disease. Adv Exp Med Biol. 2014;825:389–429.

22. The UniProt Consortium. UniProt: The universal protein knowledgebase in 2021. Nucleic Acids Res. 2021;49:D480–D489.

23. Landrum MJ, Lee JM, Benson M, et al. ClinVar: Improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46:D1062–D1067.

24. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–29.

25. Gene Ontology Consortium. The gene ontology resource: Enriching a GOld mine. Nucleic Acids Res. 2018;46:D1062–D1067.

26. McGinnis S, Madden TL. BLAST: At the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 2000;28:291–294.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Sanders AL, Hermanson JN, Samuels DC, Plate L, Sanders CR. Compendium of proteins containing segments that exhibit zero-tolerance to amino acid variation in humans. Protein Science. 2022;31(9):e4408. https://doi.org/10.1002/pro.4408