Biophysics and Physicobiology

Commentary and Perspective

Current status and future perspectives of the evaluation of missense variants by using three-dimensional structures of proteins

Matsuyuki Shirota[1,2], Kengo Kinoshita[2,3]

[1] *Graduate School of Medicine, Tohoku University, Sendai, Miyagi 980-8575, Japan*

[2] *Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi 980-8575, Japan*

[3] *Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi 980-8579, Japan*

Changes in the amino acid sequences of proteins may cause changes in molecular function, resulting in phenotypic variations among species and individuals. Such amino acid changes occur naturally due to mutations in the genome sequence of an organism and can be inherited. Recent advances in genome sequencing technologies have enabled us to sequence the genomes of millions of humans, as performed by various large-scale projects such as the Genome Aggregation Database (gnomAD) [1], Trans-Omics for Precision Medicine (TOPMed) [2], and UK Biobank [3]. Further, it has enabled us to accumulate known pathogenic variants in databases such as ClinVar [4], Human Genome Mutation Database (HGMD) [5] and Catalogue of Somatic Mutations in Cancer (COSMIC) [6]. These studies have identified a large number of genomic variants, an important part of which are missense single-nucleotide variants (SNVs), which change the amino acid sequence of proteins. Identifying whether and how these amino acid changes result in differences in human phenotypes, such as complex traits and diseases, is a major challenge in genomic medicine. In particular, rarely observed variants make it difficult to interpret their functional impacts using conventional genome-wide association studies. Biophysics can play a crucial role in this task by illuminating the mechanisms by which amino acid changes affect protein stability, function, and interaction.

To evaluate the impact of an amino acid change, three-dimensional (3D) protein structures in the Protein Data Bank (PDB) [7] provide a wealth of information. Several tools have been developed to evaluate the structural positions of amino acids using genome coordinates [8,9]. However, the accuracy of the prediction methods of protein 3D structures from amino acid sequences has achieved significant progress in recent years owing to methodological breakthroughs using protein co-evolution and deep learning algorithms. AlphaFold [10] showed remarkable prediction accuracy in the last Critical Assessment of Techniques for Protein Structure Prediction (CASP14) held in 2020. Models generated with AlphaFold for all human proteins are now used in various biological fields [11]. Progress in the coverage of the human proteome by both experimental and modeled structures allows us to interpret the impact of missense SNVs on human phenotypes through detailed inspection of the conformation of changed amino acids, although how the coverage of structure information has increased over time has not been well described.

In this Commentary and Perspective, we describe the use of experimental and modeled protein structures to interpret publicly available SNVs obtained by sequencing healthy and diseased individuals. We report the coverage of the human proteome and missense variants by 3D structures over time in comparison to the proteins covered by AlphaFold DB models. We discuss how recent protein structures have facilitated the discrimination of pathogenic and benign variants and the importance of expanding the knowledge of biological protein structures to understand the effects of missense variants.

We mapped the bases in the GRCh37 human genome to the amino acids of human proteins based on the protein-coding regions defined by the consensus coding sequence project (CCDS) version 20131129 [12]. The gene symbol for each protein coding sequence was obtained from the RefSeq database version 105 [13]. For genes having multiple isoforms in

---

Corresponding author: Kengo Kinoshita, Graduate School of Information Sciences, Tohoku University, 6-3-09, Aza-Aoba, Aoba-ku, Sendai, Miyagi 980-8579, Japan. ORCID iD: https://orcid.org/0000-0003-3453-2171, e-mail: kengo@tohoku.ac.jp

CCDS, we prioritized the isoforms to select the representative one in the following criteria; 1) the Select isoform, which was selected by NCBI or MANE (Matched Annotation between NCBI and EBI-EMBL) [14], 2) the isoform with the same amino acid sequence as the Select isoform, 3) the isoform having the amino acid sequence of the same length with the Select isoform but with at most 10 amino acid mismatches and 4) the one with the longest amino acid sequence. We extracted the amino acid sequences of all polypeptide (L) entities from all PDB entries as of April 6th, 2022. Next, we used BLAST to search these PDB entities for human protein sequences at a sequence identity threshold of 30% and then performed pairwise alignment of the significant hit sequences. If an amino acid of a human protein was aligned to that in the PDB with coordinate information (i.e., excluding disordered residues), it was considered to be covered in PDB. Gene symbols were obtained from RefSeq for human proteins and from UniProt for PDB proteins. If the human protein and aligned PDB protein were coded by the same gene, the amino acid was covered by identity; otherwise, it was covered by homology. In addition, if at least 100 residues or half of the protein residues were covered by the PDB, the protein was considered to be covered. We restricted the PDB entries to those published before April 6th, 2022, and evaluated how the coverage of human proteins and amino acids by identity or homology has been increasing with time. We also obtained structural models of human proteins from the AlphaFold DB, aligned the amino acid sequences of RefSeq and that of the AlphaFold DB for each gene, and evaluated the coverage of the human proteins and amino acids that were mapped to AlphaFold models. In addition, we also calculated the coverage in AlphaFold DB using the residues with a predicted local distance difference test [15] (pLDDT) score > 50, which can be regarded as predicted as ordered residues [10]. It should be noted that discrepancies between protein-based and genome-based databases are problematic in determining the correspondence between amino acid changes and genome missense SNVs, and we followed CCDS mapping in this study. See Shirota and Kinoshita [16] for details on the discrepancies between the databases.

As of April 6th, 2022, 32.9% and 68.3% of human proteins were covered by identity and homology, respectively, whereas 95.8% and 94.8% were covered by all and predicted ordered (pLDDT > 50) residues of AlphaFold DB, respectively (Figure 1A and Supplementary Table S1A). Despite the steady increase in PDB coverage, the availability of experimental structures is still modest, and models predicted by AlphaFold DB can compensate for missing structure information. In contrast, 19.2% and 42.7% of human amino acids were covered by identity and homology, respectively (Figure 1B and Supplementary Table S1B). The smaller coverage of amino acids than that of proteins is attributed to the existence of disordered residues, for which structure determination is difficult to obtain. The coverage of amino acids by all of the residues in AlphaFold DB was 94.9%, whereas that by predicted ordered residues (pLDDT>50) in AlphaFold DB was 69.6%, indicating that approximately 25% of the human amino acids were predicted to be disordered.
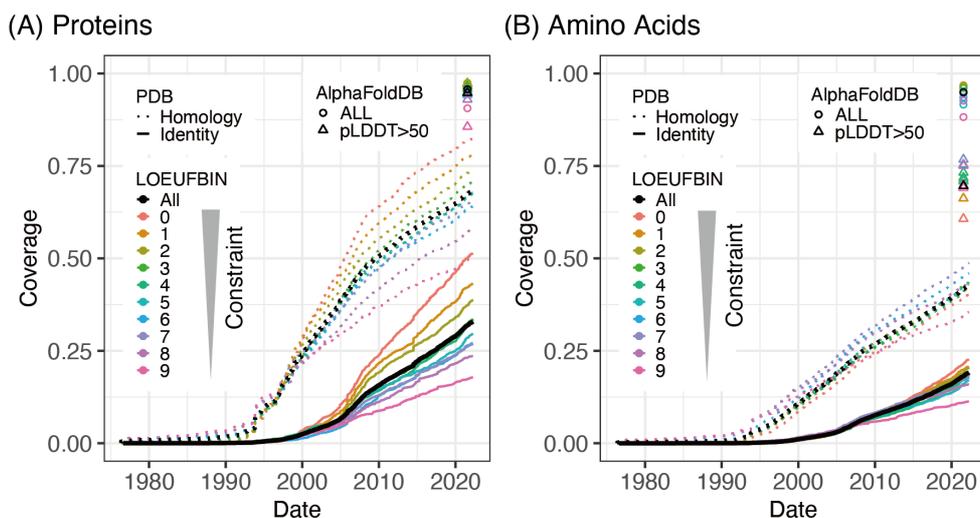

Figure 1 Accumulative coverage of human (A) proteins and (B) amino acids by the structures in the PDB as function of time. Solid and black lines indicate the coverage by identity and homology, respectively. Black lines show the coverage over the entire human proteins and colored lines show the coverage in the subsets of the proteins classified by the LOEUF bins of their genes. Circles and triangles show the coverages by all of the residues and by the residues with pLDDT > 50, respectively, in the structures of AlphaFold DB.

We then divided human proteins into 10 bins using the loss-of-function observed/expected upper-bound fraction (LOEUF) scores [1] of the corresponding genes. A smaller LOEUF bin indicates stronger constraints on genes such that fewer protein-truncating variants, including nonsense, frameshift, or canonical splice site variants, are observed for the genes in the bin among a population. Protein coverage, both by identity and homology, correlated with gene constraint,

thus indicating that more constrained genes are important targets of structural biology (Figure 1A, Supplementary Figure S1A). The relationship between gene constraint and amino acid coverage by identity was not as clear as that with protein coverage due to different fractions of predicted disordered residues of the proteins in each LOEUF bin (Figure 1B, Supplementary Figure S1B).

The rate of increase in protein and amino acid coverage by identity first increased around 2005 and 2007, then slowed between 2008 and 2013, and accelerated after 2014. On the other hand, the rate of increase in homology was high until 2007 but slowed down after 2008. These trends may indicate that recent structural determination efforts have focused on human proteins rather than on their homologues.

To elucidate how different technologies contributed to solving new structures of human proteins over time, we counted the number of amino acids covered in PDB for the first time each year and classified them using the experimental method used for structure determination (Figure 2). X-ray crystallography (XRAY) was used for the majority of the new amino acids until 2019, ranging from approximately 50,000 to 80,000 amino acids per year since 2000. NMR was used to identify new human proteins that account for approximately 25,000 amino acids per year between 2005 and 2007 when the contribution of NMR to new amino acids culminated. This increase in both XRAY and NMR structure determination resulted in a rapid increase in protein and amino acid coverage (Figure 1). However, the number of new amino acids covered by these methods decreased after 2008, which led to a slowed increase in protein and amino acid coverage between 2008 and 2013. The third method, cryogenic electron microscopy (EM), has rapidly evolved since 2014 and exceeded XRAY in 2020 with respect to the number of new amino acids covered; this has boosted the coverage by identity since 2014 (Figure 1). In 2021, most of the new amino acids were solved using EM. These trends suggest that innovations in structure determination technologies are needed for the unsolved human proteins.
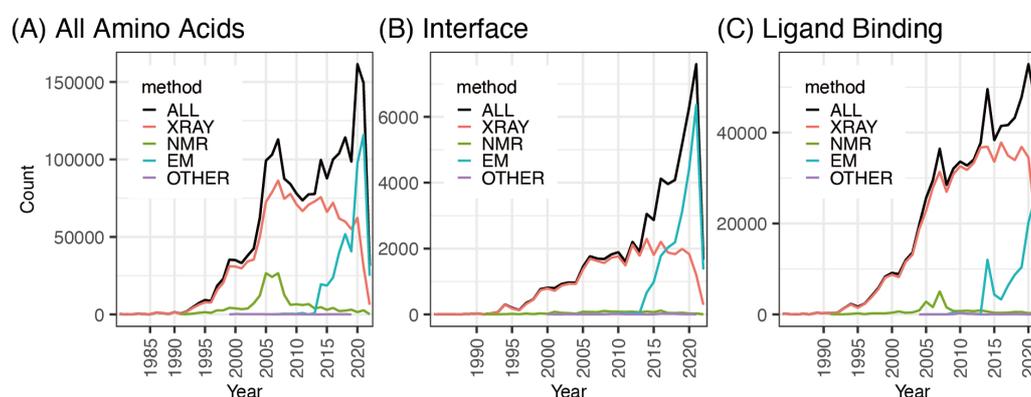

Figure 2 The number of human amino acids observed for the first time in PDB structures each year. (A) all amino acids, (B) the amino acids in protein-protein interfaces and (C) the amino acids in ligand binding sites. The amino acids in interfaces are defined by relative accessible surface area (ASA) in complex < 0.25 and the difference in ASA (ΔASA) between in the monomeric and complex states > 0.5. The ligand binding residues were defined if the minimum inter-atomic distance to any atom of non-protein and non-water molecules is smaller than $10\,\text{Å}$. The counts for 2022 are decreased because they are calculated with the structures published as of April 6th.

We then focused on the amino acids that were found for the first time in each year to be located in two types of functional regions: protein–protein interaction interfaces (Figure 2B) and ligand-binding sites (Figure 2C). The number of interface amino acids was smaller than the total number of amino acids. These amino acids were almost exclusively found with XRAY before 2013, whereas since 2014, they have been found increasingly with EM, thus contributing to a three-fold increase in number in 2021. NMR contributed little to the identification of interface amino acids, probably because it was used to analyze the monomeric structures of small proteins. On the other hand, most of the new amino acids in ligand-binding sites were identified with XRAY, while the number solved by Cryo-EM exceeded that solved by XRAY in 2021. EM has been used to solve large protein-protein and large protein-nucleic acid complexes, which include a large area of protein-protein interaction sites and a large number of ligand (e.g., DNA or RNA) binding sites. For example, the structure of the human small-subunit processome [PDB ID: 7MQ8, 7MQ9, 7MQA] includes 20,341 new amino acids, of which 489 are in interfaces and 5,099 are in ligand-binding sites, while the EM structure of the human RNA polymerase I complex [PDB ID: 7OB9, 7OBA, 7OBB] includes 3,882 new amino acids, of which 220 are in interfaces and 585 are in ligand sites (Supplementary Figure S2). The ability of EM to solve large complexes allowed us to analyze a much larger variety and number of protein functional sites, which cannot be obtained from monomeric AlphaFold DB models.

To evaluate the effects of these new amino acids on the interpretation of missense variants in healthy and diseased

populations, we counted the number of human amino acids that were changed by a certain class of SNVs. The following variants were considered: gnomAD variants were obtained from the gnomAD website. They were classified as either singleton (allele count = 1), low-frequency (allele count > 1 and allele frequency < 0.01), or common (allele frequency > 0.01). Variants related to the disease were downloaded from ClinVar. They were classified as pathogenic (CLNSIG = pathogenic or likely pathogenic), VUS (CLNSIG = uncertain significance), or benign (CLNSIG = benign or likely benign). The variants found in patients with cancer were downloaded from COSMIC.

The number of new amino acids with variants covered in PDB each year increased until 2005 and reached a plateau after 2005. In the 2010s, the number of amino acids covered each year was > 10,000 for COSMIC, gnomAD singleton, and gnomAD low-frequency variants, several thousand for ClinVar VUS, ~1,000 for ClinVar pathogenic variants, and several hundred for ClinVar benign and gnomAD common variants (Figure 3). Each year, structural information is available for approximately 1,000 pathogenic variants. Such structures with known pathogenic variants can be a good resource for inferring the biological effects of a much larger number of VUSs, for which pathogenicity is unclear, for the same protein. Experimental structures can also be used to predict variants that have deleterious effects on protein function from variants found in the general population such as gnomAD. Such variants with strong effects are usually rare (low frequency or singletons), which makes statistical testing of their effect on phenotypes with genome-wide association studies difficult. Biophysical analysis, both computational and experimental, can help interpret these.
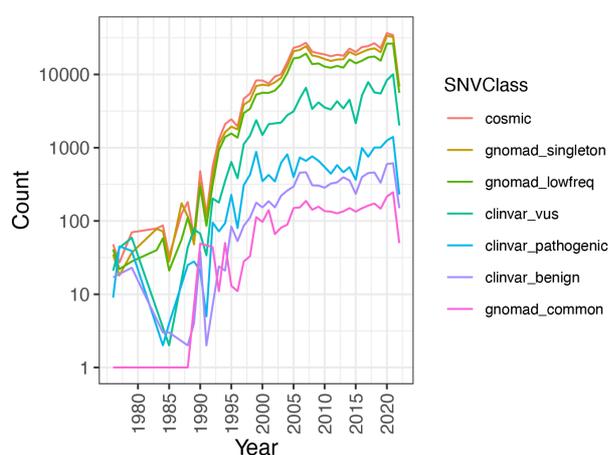

Figure 3 The number of human amino acids with known missense variants that were for the first time covered in PDB each year.

We then counted the number of amino acids that were covered for the first time between January 6th, 2021, and April 6th, 2022, which were changed by known ClinVar pathogenic variants (Figure 4). The number of amino acids with pathogenic variants covered for the first time by new PDB entries varied each day, depending on whether clinically important protein structures were solved. The number of amino acids with pathogenic variants also varied among the genes. The genes for which large number of amino acids with pathogenic variants are covered in PDB during this period are *SCN1A* with 303 amino acids (on April 7th, 2021), *ABCA4* with 182 amino acids (on March 3rd, 2021), and *NF1* with 122 amino acids (November 11th, 2021). We compared the amino acids with population and pathogenic variants in the 3D structures of *SCN1A* and *NF1*.

NF1 is the causal gene of neurofibromatosis type I, an autosomal dominant disease characterized by skin lesions (café-au-lait spots) and tumors of neurons called neurofibroma [17]. *NF1* encodes the neurofibromin (NF1) protein, which binds to the Ras protein and inhibits its
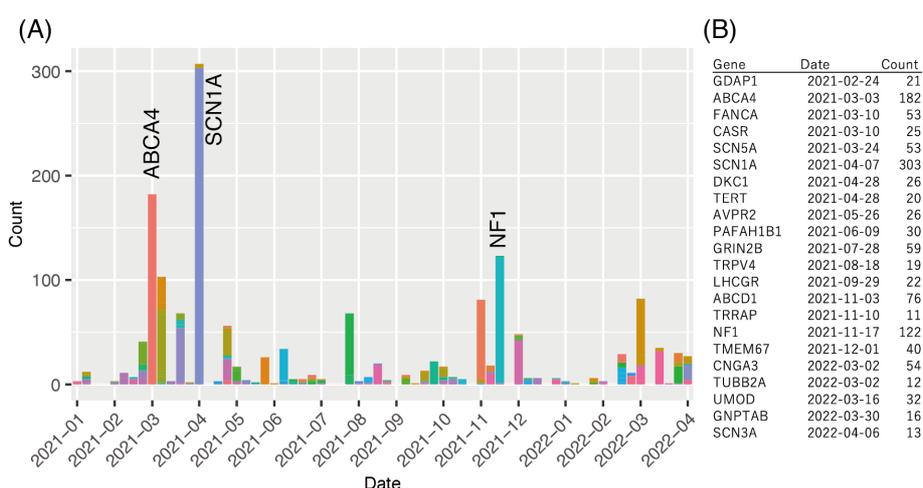

Gene	Date	Count
GDAP1	2021-02-24	21
ABCA4	2021-03-03	182
FANCA	2021-03-10	53
CASR	2021-03-10	25
SCN5A	2021-03-24	53
SCN1A	2021-04-07	303
DKC1	2021-04-28	26
TERT	2021-04-28	20
AVPR2	2021-05-26	26
PAFAH1B1	2021-06-09	30
GRIN2B	2021-07-28	59
TRPV4	2021-08-18	19
LHCGR	2021-09-29	22
ABCD1	2021-11-03	76
TRRAP	2021-11-10	11
NF1	2021-11-17	122
TMEM67	2021-12-01	40
CNGA3	2022-03-02	54
TUBB2A	2022-03-02	12
UMOD	2022-03-16	32
GNPTAB	2022-03-30	16
SCN3A	2022-04-06	13

Figure 4 (A) The number of amino acids that were covered in PDB for the first time each day and that have known pathogenic variants in ClinVar. (B) The genes with at least 10 amino acids that were covered in PDB for the first time each day and that have known pathogenic variants in ClinVar.
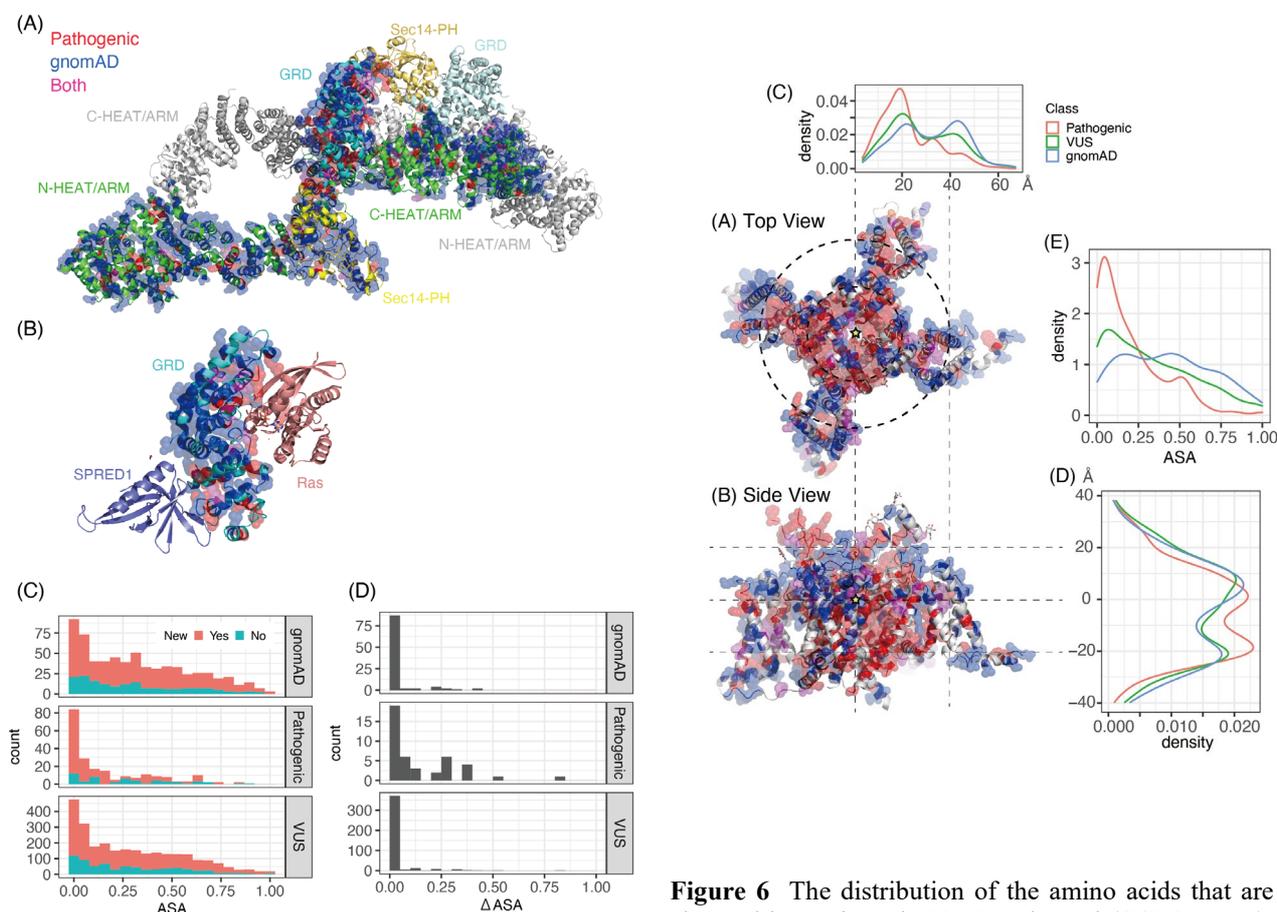
Figure 5 (A) EM structure [PDB ID: 7PGT] of a dimeric NF1 protein. N-HEAT/ARM, GRD, Sec14-PH and C-HEAT/ARM domains are shown in different colors. The amino acids with ClinVar pathogenic, with gnomAD variants or with both variants are shown in red, blue and magenta spheres, respectively. (B) X-ray structure [PDB ID: 6V65] of a complex of NF1 GRD, Ras and SPRED1. The amino acids changed by variants are colored as in (A). (C) The distribution of ASA of the amino acids that are changed by missense variants in PDB 7PGT. The variants are classified if the amino acids changed by them were for the first time observed in PDB 7PGT (New is Yes), i.e. in N- and C-HEAT/ARM domains, or they were found in older PDB structures (New is No), i.e. in GRD or Sec14-PH domains. (D). The distribution of ΔASA of the amino acids that are changed by missense variants in PDB 6V65.

Figure 6 The distribution of the amino acids that are changed by variants in Nav1.1 channel (*SCN1A* gene). (A), (B) Structure of Nav1.1 [PDB ID: 7DTD] seen from the extracellular space (A) and from the membrane plane (B). The center of the Cα atoms of residues 381, 951, 1431 and 1724 in the selectivity filter is marked by a star and set as the origin of the structure. The membrane normal was defined by the vector connecting the origin and the center of the Cα atoms of residues 394, 961, 1443 and 1740. The amino acids with ClinVar pathogenic, with gnomAD variants or with both variants are shown in red, blue and magenta spheres, respectively. (C) The distributions of the radial distance of the variants from the membrane normal. (D) The distribution of the variants along the membrane normal. The region between -20 to 0 Å corresponds to the trasmembrane region. (E) The distribution of ASA of the variants. The densities in (C)-(E) are normalized for each type of variants, such that the area under each curve is equal to 1.

function, leading to tumor suppression. NF1 consists of a GTPase-activating protein-related domain (GRD) that binds to Ras, a Sec14-PH domain that binds to phospholipids, and terminal N-HEAT/ARM and C-HEAT/ARM domains. Structure of GRD in isolation [PDB ID: 1NF1] and in complex with Ras and Sprouty-related proteins with an EVT domain (SPRED1) [PDB ID: 6V65] [18] and the Sec-PH domain [PDB ID: 2D4Q] [19] have been reported, whereas the full-length structure was solved on November 17th, 2021 [PDB ID: 7PGT] in dimeric forms [20]. The full-length structure has coordinates of 154 residues with ClinVar pathogenic variants, among which 122 have not been observed in previous NF1 structures in PDB. The dimeric full-length structure of NF1 showed that pathogenic variants are almost always buried,

whereas gnomAD variants are located in exposed regions (Figure 5A and C), which is consistent with the idea that deleterious variants of the NF1 internal region lead to protein dysfunction and loss of tumor-suppression activities. In contrast, the fractions of the residues in the dimeric interface were not significantly different between the variants, although the fraction of pathogenic variants were smaller than those of gnomAD variants, indicating that dimeric interactions were not impaired in NF1 pathogenesis (Supplementary Figure S3A). However, the heterotrimeric structures of NF1 GRD, Ras, and SPRED1 [PDB ID: 6V65] showed that pathogenic variants in the GRD were mostly located at the interfaces with Ras or SPRED1 (Figure 5B and D, Supplementary Figure S3B), indicating that the interaction with Ras or SPRED1 is critical for NF1 function. NF1 VUS missense variants have widespread distribution of residue burial and inclusion at the interfaces. These structural features can be used to filter pathogenic variants in VUS.

SCN1A encodes the α-subunit of the voltage-gated sodium channel Nav1.1, which is expressed in the brain and whose mutations have been reported in patients with epilepsy [21]. The EM structure of Nav1.1, published on April 7th, 2021 [PDB ID: 7DTD] [22], includes the coordinates of 303 residues that are mutated by ClinVar pathogenic variants. Nav1.1 α-subunit forms a pseudo-tetrameric conformation and pathogenic variants are clustered in the pore domain located in the center, whereas four voltage-sensing domains in the peripheral are also hotspots of pathogenic mutations (Figure 6A-C). In contrast, gnomAD variants were observed mostly away from the pore. Pathogenic variants were also located in the transmembrane regions, whereas gnomAD variants were observed in the cytosolic and extracellular regions (Figure 6D). The densities have two peaks at the membrane-cytosol and membrane-extracellular borders, which can be interpreted as a suppression of missense variants in the inner-membrane region due to their harm to protein function. The amino acids changed by pathogenic variants were more frequently found in the buried regions of Nav1.1 than those changed by gnomAD variants (Figure 6E). These structural features are also important for estimating the effect of VUSs, which show intermediate distributions between the pathogenic and gnomAD variants.

In this study, we evaluated how the coverage of human proteins and amino acids in the PDB grew over time. We observed that important proteins with strong constraints have been intensively studied in the field of structural genomics with accelerated speed fueled by methodological breakthroughs and international initiatives. Predicted models with AlphaFold have become a common tool for interpreting missense variants, covering almost the entire human proteome. Though we focused on structured protein regions, it should be noted that missense variants in the disordered regions, which were predicted to occupy 25% of the human amino acids, can affect protein function because they frequently include important residues, which can be chemically modified by reactions such as phosphorylation, methylation, or acetylation. Experimental structures in PDB are not as comprehensively available as AlphaFold models but have the advantage of providing the involvement of residues in protein-protein and protein-ligand interactions in biologically active protein complexes. New insights were obtained on the functional consequences of previously determined missense variants by weekly updating of PDB structures for reanalysis of variant annotation. These results highlight the importance of new protein structures in bridging the gap between human genotypes and phenotypes.

References

[1] Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434-443 (2020). https://doi.org/10.1038/s41586-020-2308-7

[2] Taliun, D., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., Torres, R., et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature 590, 290-299 (2021). https://doi.org/10.1038/s41586-021-03205-y

[3] Backman, J. D., Li, A. H., Marcketta, A., Sun, D., Mbatchou, J., Kessler, M. D., et al. Exome sequencing and analysis of 454,787 UK Biobank participants. Nature 599, 628-634 (2021). https://doi.org/10.1038/s41586-021-04103-z

[4] Landrum, M. J., Chitipiralla, S., Brown, G. R., Chen, C., Gu, B., Hart, J., et al. ClinVar: Improvements to accessing data. Nucleic Acids Res. 48, D835-D844 (2020). https://doi.org/10.1093/nar/gkz972

[5] Stenson, P. D., Mort, M., Ball, E. V., Evans, K., Hayden, M., Heywood, S., et al. The Human Gene Mutation Database: Towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Hum. Genetics 136, 665-677 (2017). https://doi.org/10.1007/s00439-017-1779-6

[6] Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., et al. COSMIC: The catalogue of somatic mutations in cancer. Nucleic Acids Res. 47, D941-D947 (2019). https://doi.org/10.1093/nar/gky1015

[7] wwPDB consortium. The Protein Data Bank: The single global archive for 3D macromolecular structure data. Nucleic Acids Res. 47, D520-D528 (2019). https://doi.org/10.1093/nar/gky949

[8] Stephenson, J. D., Laskowski, R. A., Nightingale, A., Hurles, M. E., Thornton, J. M. VarMap: A web tool for mapping genomic coordinates to protein sequence and structure and retrieving protein structural annotations. Bioinformatics 35, 4854-4856 (2019). https://doi.org/10.1093/bioinformatics/btz482

[9] Niknafs, N., Kim, D., Kim, R., Diekhans, M., Ryan, M., Stenson, P. D., et al. MuPIT interactive: Webserver for mapping variant positions to annotated, interactive 3D structures. Human Genetics 132, 1235-1243 (2013). https://doi.org/10.1007/s00439-013-1325-0

[10] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583-589 (2021). https://doi.org/10.1038/s41586-021-03819-2

[11] Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Zidek, A., et al. Highly accurate protein structure prediction for the human proteome. Nature 596, 590-596 (2021). https://doi.org/10.1038/s41586-021-03828-1

[12] Farrell, C. M., O'Leary, N. A., Harte, R. A., Loveland, J. E., Wilming, L. G., Wallin, C., et al. Current status and new features of the Consensus Coding Sequence database. Nucleic Acids Res. 42, D865-D872 (2014). https://doi.org/10.1093/nar/gkt1059

[13] Pruitt, K. D., Brown, G. R., Hiatt, S. M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., et al. RefSeq: An update on mammalian reference sequences. Nucleic Acids Res. 42, D756-D763 (2014). https://doi.org/10.1093/nar/gkt1114

[14] Morales, J., Pujar, S., Loveland, J. E., Astashyn, A., Bennett, R., Berry, A., et al. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. Nature 604, 310-315 (2022). https://doi.org/10.1038/s41586-022-04558-8

[15] Mariani, V., Biasini, M., Barbato, A., Schwede, T. lDDT: A local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics 29, 2722-2728 (2013). https://doi.org/10.1093/bioinformatics/btt473

[16] Shirota, M., Kinoshita, K. Discrepancies between human DNA, mRNA and protein reference sequences and their relation to single nucleotide variants in the human population. Database: The journal of biological databases and curation 2016, baw124 (2016). https://doi.org/10.1093/database/baw124

[17] Hirbe, A. C., Gutmann, D. H. Neurofibromatosis type 1: A multidisciplinary approach to care. Lancet Neurol. 13, 834-843 (2014). https://doi.org/10.1016/s1474-4422(14)70063-8

[18] Yan, W., Markegard, E., Dharmaiah, S., Urisman, A., Drew, M., Esposito, D., et al. Structural insights into the SPRED1-Neurofibromin-KRAS complex and disruption of SPRED1-Neurofibromin interaction by oncogenic EGFR. Cell Rep. 32, 107909 (2020). https://doi.org/10.1016/j.celrep.2020.107909

[19] D'Angelo, I., Welti, S., Bonneau, F., Scheffzek, K. A novel bipartite phospholipid-binding module in the neurofibromatosis type 1 protein. EMBO Rep. 7, 174-179 (2006). https://doi.org/10.1038/sj.embor.7400602

[20] Naschberger, A., Baradaran, R., Rupp, B., Carroni, M. The structure of neurofibromin isoform 2 reveals different functional states. Nature 599, 315-319 (2021). https://doi.org/10.1038/s41586-021-04024-x

[21] Scheffer, I. E., Nabbout, R. SCN1A-related phenotypes: Epilepsy and beyond. Epilepsia 60 Suppl 3, S17-S24 (2019). https://doi.org/10.1111/epi.16386

[22] Pan, X., Li, Z., Jin, X., Zhao, Y., Huang, G., Huang, X., et al. Comparative structural analysis of human Nav1.1 and Nav1.5 reveals mutational hotspots for sodium channelopathies. Proc. Natl. Acad. Sci. U.S.A. 118, e2100066118 (2021). https://doi.org/10.1073/pnas.2100066118