

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
20 October 2022 (20.10.2022)



(10) International Publication Number
WO 2022/221591 A1

(51) International Patent Classification:
G16B 15/00 (2019.01) *G16B 40/20* (2019.01)
G16B 20/00 (2019.01)

(21) International Application Number:
PCT/US2022/024916

(22) International Filing Date:
14 April 2022 (14.04.2022)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/175,495 15 April 2021 (15.04.2021) US
63/175,767 16 April 2021 (16.04.2021) US
17/703,935 24 March 2022 (24.03.2022) US
17/703,958 24 March 2022 (24.03.2022) US

(71) Applicants: **ILLUMINA, INC.** [US/US]; 5200 Illumina Way, San Diego, California 92122 (US). **ILLUMINA CAMBRIDGE LIMITED** [GB/GB]; 19 Granta Park, Great Abington, Cambridge CB21 6DF (GB).

(72) Inventors: **HAMP, Tobias**; 19 Granta Park, Great Abington, Cambridge CB21 6DF (GB). **GAO, Hong**; 5200 Illumina Way, San Diego, California 92122 (US). **FARH, Kai-How**; 5200 Illumina Way, San Diego, California 92122 (US).

(74) Agent: **KHAN, Sikander M.** et al.; Haynes Beffel & Wolfeld LLP, P.O. Box 366, 637 Main Street, Half Moon Bay, California 94019 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,

(54) Title: MULTI-CHANNEL PROTEIN VOXELIZATION TO PREDICT VARIANT PATHOGENICITY USING DEEP CONVOLUTIONAL NEURAL NETWORKS

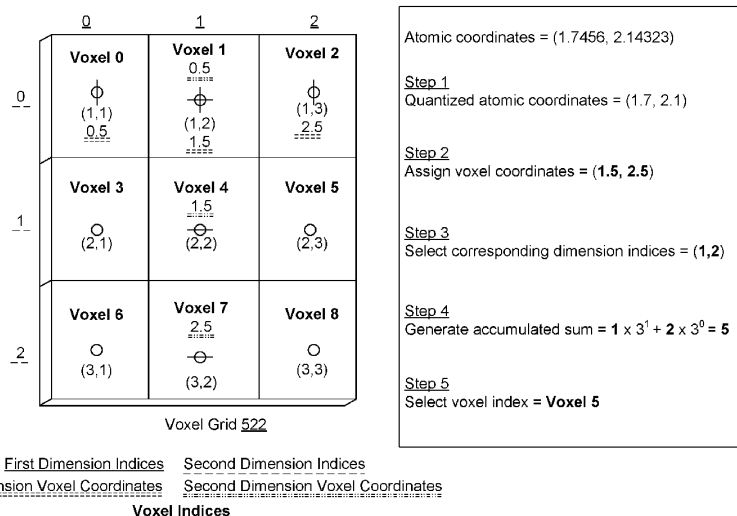


Figure 33

(57) Abstract: A system includes at least a voxelizer, an alternative allele encoder, an evolutionary conservation encoder, and a convolutional neural network. The voxelizer accesses a three-dimensional structure of a reference amino acid sequence of a protein and fits a three-dimensional grid of voxels on atoms in the three-dimensional structure on an amino acid-basis to generate amino acid-wise distance channels. The alternative allele encoder encodes an alternative allele sequence to each voxel in the three-dimensional grid of voxels. The evolutionary conservation encoder encodes an evolutionary conservation sequence to each voxel in the three-dimensional grid of voxels. The convolutional neural network applies three-dimensional convolutions to a tensor that includes the amino acid-wise distance channels encoded with the alternative allele sequence and respective evolutionary conservation sequences and determines a pathogenicity of a variant nucleotide based at least in part on the tensor.

[Continued on next page]

DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

- *with international search report (Art. 21(3))*
-