

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
21 October 2021 (21.10.2021)



(10) International Publication Number
WO 2021/211840 A1

(51) International Patent Classification:

C12P 19/34 (2006.01) C40B 30/02 (2006.01)
C12Q 1/68 (2018.01) G06F 19/22 (2011.01)
C12Q 1/6881 (2018.01)

(21) International Application Number:

PCT/US2021/027478

(22) International Filing Date:

15 April 2021 (15.04.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/010,467 15 April 2020 (15.04.2020) US

(71) Applicants: CHAN ZUCKERBERG BIOHUB, INC.

[US/US]; 499 Illinois Street, San Francisco, California 94158 (US). **THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY** [US/US]; Office of the General Counsel, Bldg. 170, Third Floor, Main Quad, P.O. Box 20386, Stanford, California 94305-2038 (US).

(72) Inventors: MONTSERRAT, Daniel Mas; c/o The Board

of Trustees of the Leland Stanford Junior University, Office of the General Counsel, Bldg. 170, Third Floor, Main

Quad, P.O. Box 20386, Stanford, California 94305-2038 (US). **IOANNIDIS, Alexander**; c/o The Board of Trustees of the Leland Stanford Junior University, Office of the General Counsel, Bldg. 170, Third Floor, Main Quad, P.O. Box 20386, Stanford, California 94305-2038 (US). **KUMAR, Arvind**; c/o The Board of Trustees of the Leland Stanford Junior University, Office of the General Counsel, Bldg. 170, Third Floor, Main Quad, P.O. Box 20386, Stanford, California 94305-2038 (US). **BUSTAMANTE, Carlos**; c/o The Board of Trustees of the Leland Stanford Junior University, Office of the General Counsel, Bldg. 170, Third Floor, Main Quad, P.O. Box 20386, Stanford, California 94305-2038 (US). **RASTOGI, Richa**; c/o The Board of Trustees of the Leland Stanford Junior University, Office of the General Counsel, Bldg. 170, Third Floor, Main Quad, P.O. Box 20386, Stanford, California 94305-2038 (US). **HILMARSSON, Helgi**; c/o The Board of Trustees of the Leland Stanford Junior University, Office of the General Counsel, Bldg. 170, Third Floor, Main Quad, P.O. Box 20386, Stanford, California 94305-2038 (US).

(74) Agent: CHOY, Ming Wai et al.; Kilpatrick Townsend & Stockton LLP, Mailstop: IP Docketing - 22, 1100 Peachtree Street, Suite 2800, Atlanta, Georgia 30309 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

(54) Title: LOCAL-ANCESTRY INFERENCE WITH MACHINE LEARNING MODEL

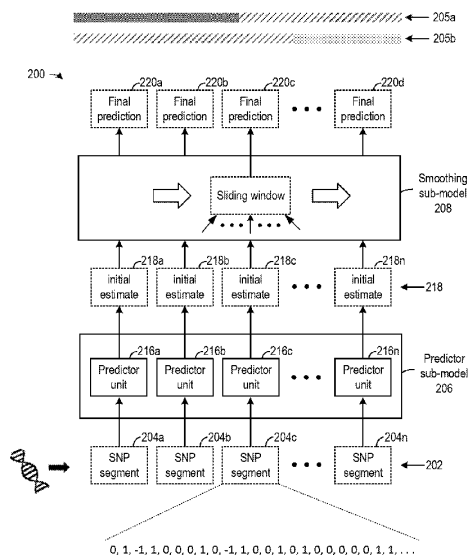


FIG. 2A

(57) Abstract: A computer-implemented method comprises: storing a trained machine learning model, the machine learning model comprising a predictor sub-model and a smoothing sub-model, the machine learning model being trained based on segments of training genomic sequences that have known ancestral origins; receiving data representing an input genomic sequence of the subject, the input genomic sequence covering a plurality of segments including a plurality of single nucleotide polymorphisms (SNP) sites of the genome of the subject, wherein each segment comprises a sequence of SNP values at the SNP sites, each SNP value specifying a variant at the SNP site; determining, using the predictor sub-model and based on the data, an initial ancestral origin estimate of each segment of SNP values; and performing, by the smoothing sub-model for each segment, a smoothing operation over the initial ancestral origin estimates to obtain a final prediction result for the ancestral origin of the segment.

[Continued on next page]

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, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, 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).

Published:

— with international search report (Art. 21(3))