

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

(19) World Intellectual Property
Organization
International Bureau

(43) International Publication Date
11 April 2024 (11.04.2024)



(10) International Publication Number
WO 2024/077096 A1

(51) International Patent Classification:

G16B 20/20 (2019.01) G16B 30/00 (2019.01)
C12Q 1/6869 (2018.01) G16B 40/20 (2019.01)

(21) International Application Number:

PCT/US2023/075999

(22) International Filing Date:

04 October 2023 (04.10.2023)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/378,474 05 October 2022 (05.10.2022) US
63/482,163 30 January 2023 (30.01.2023) US

(71) Applicant: ILLUMINA, INC. [US/US]; 5200 Illumina Way, San Diego, CA 92122 (US).

(72) Inventors: PARNABY, Gavin Derek; 5200 Illumina Way, San Diego, California 92122 (US). HASHEMIDOULABI, Seyedmohammadjafar; 5200 Illumina Way, San Diego, California 92122 (US). HALPERN, Aaron L.; 5200 Illumina Way, San Diego, California 92122 (US). RUEHLE, Michael; 5200 Illumina Way, San Diego, California 92122 (US).

(74) Agent: FUNK, Christopher C. et al.; KELLER PREECE PLLC, 1010 North 500 East, Suite 210, North Salt Lake, Utah 84054 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY,

(54) Title: INTEGRATING VARIANT CALLS FROM MULTIPLE SEQUENCING PIPELINES UTILIZING A MACHINE LEARNING ARCHITECTURE

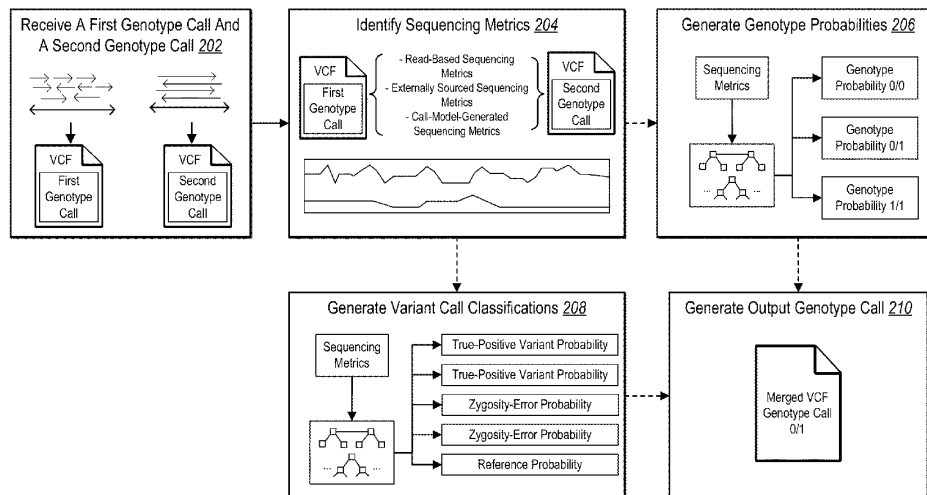


Fig. 2

(57) Abstract: This disclosure describes methods, non-transitory computer readable media, and systems that can generate genotype calls from a combined pipeline for processing nucleotide reads from multiple read types/sources for robust, accurate genotype calls. For example, the disclosed systems can train and/or utilize a genotype-call-integration machine-learning model to generate predictions for genotype calls based on data associated with a first type of nucleotide reads (e.g., short reads) and a second type of nucleotide reads (e.g., long reads). As disclosed, the disclosed systems can determine sequencing metrics and can utilize a genotype-call-integration machine-learning model to generate predictions (e.g., genotype probabilities, variant call classifications) for generating output genotype calls based on the sequencing metrics. The disclosed system can utilize multiple such genotype-call-integration machine-learning models to generate genotype calls for different variant types, such as SNPs and indels, where the genotype-call-integration machine-learning

[Continued on next page]

MA, MD, MG, MK, MN, MU, 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, CV,
GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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))*
- *before the expiration of the time limit for amending the
claims and to be republished in the event of receipt of
amendments (Rule 48.2(h))*

models generate different predictions for each variant type.