



- (51) **International Patent Classification:**
G16B 30/00 (2019.01) *C12Q 1/6869* (2018.01)
G16B 5/20 (2019.01)
- (21) **International Application Number:**
PCT/US2023/064955
- (22) **International Filing Date:**
24 March 2023 (24.03.2023)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
63/323,995 25 March 2022 (25.03.2022) US
- (71) **Applicants:** **ILLUMINA, INC.** [US/US]; 5200 Illumina Way, San Diego, California 92122 (US). **ILLUMINA SOFTWARE, INC.** [US/US]; 5200 Illumina Way, San Diego, California 92122 (US).
- (72) **Inventors:** **VESSERE, Gery**; 5200 Illumina Way, San Diego, California 92122 (US). **DUTTA, Anindita**; 5200 Illumina Way, San Diego, California 92122 (US). **PARNABY, Gavin Derek**; 5200 Illumina Way, San Diego, California 92122 (US).
- (74) **Agent:** **FUNK, Christopher** 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, 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,

(54) **Title:** SEQUENCE-TO-SEQUENCE BASE CALLING

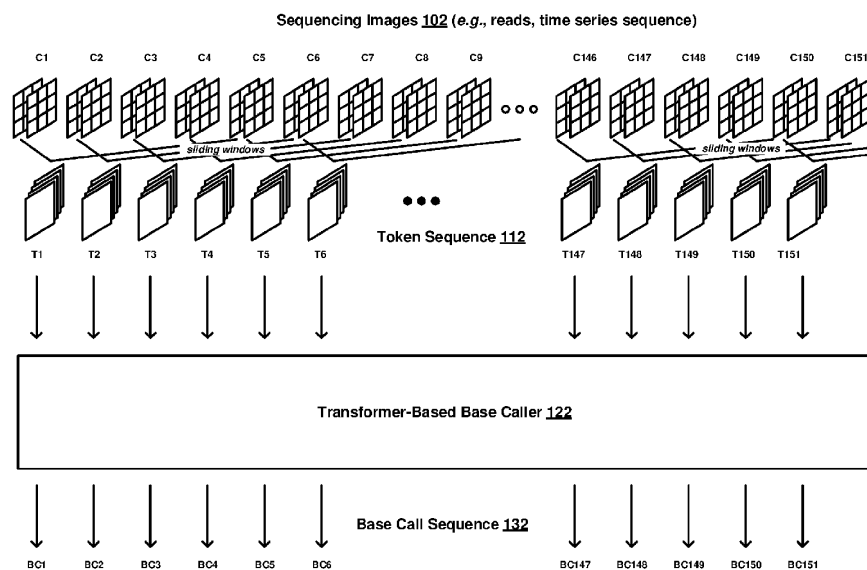


Figure 1

(57) **Abstract:** We disclose a computer-implemented method of base calling. The technology disclosed accesses a time series sequence of a read. Respective time series elements in the time series sequence represent respective bases in the read. Then, a composite sequence for the read is generated based on respective aggregate transformations of respective sliding windows of time series elements in the time series sequence. A subject composite element in the composite sequence is generated based on an aggregate transformation of a corresponding window of time series elements in the time series sequence. Then, the composite sequence is processed as an aggregate and generates a base call sequence that has respective base calls for the respective bases in the read.

[Continued on next page]

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))