



- (51) **International Patent Classification:**
G16B 20/00 (2019.01) *G16B 20/20* (2019.01)
- (21) **International Application Number:**
PCT/US2020/021550
- (22) **International Filing Date:**
06 March 2020 (06.03.2020)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
62/815,322 07 March 2019 (07.03.2019) US
- (71) **Applicant:** **ILLUMINA, INC.** [US/US]; 5200 Illumina Way, San Diego, California 92122 (US).
- (72) **Inventors:** **DOLZHENKO, Egor**, 5200 Illumina Way, San Diego, California 92122 (US). **EBERLE, Michael A.**, 5200 Illumina Way, San Diego, California 92122 (US).
- (74) **Agent:** **TSAI, Patricia** et al.; Weaver Austin Villeneuve & Sampson LLP, P. O. BOX 70250, Oakland, California 94612-0250 (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, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, 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))*

- (54) **Title:** SEQUENCE-GRAPH BASED TOOL FOR DETERMINING VARIATION IN SHORT TANDEM REPEAT REGIONS

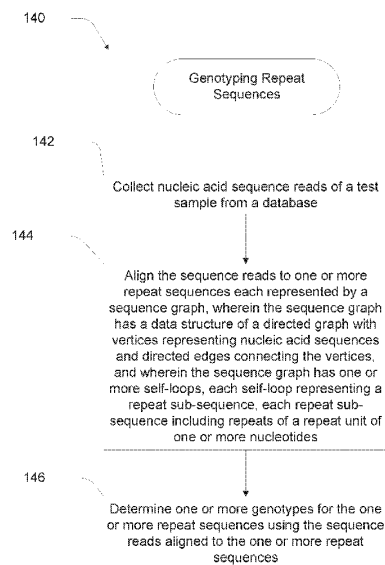


Figure 1C

(57) **Abstract:** The disclosed embodiments concern methods, apparatus, systems and computer program products for genotyping repeat sequences such as medically significant short tandem repeats (STRs). The methods involve aligning reads to a repeat sequence represented by a sequence graph, and using the aligned reads to genotype the repeat sequence. The sequence graph is a directed graph each including at least one self-loop representing a repeat sub-sequence. In some implementations, the reads are paired end reads, and both mates of each read pair may be used to genotype the repeat sequences. Some implementations can be used to determine degenerate codon repeats. Some implementations can be used to genotype repeat sequences each including two or more repeat sub-sequences. Some implementations can be used to genotype nucleic acid sequences each including at least one repeat sub-sequence and another genetic variant such as an insertion, deletion, or substitution.

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))