



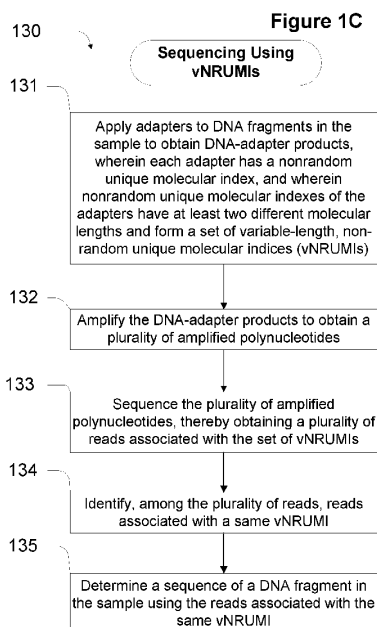
- (51) **International Patent Classification:**
G06F 19/22 (2011.01) *G06F 19/20* (2011.01)
- (21) **International Application Number:**
PCT/US2018/012669
- (22) **International Filing Date:**
05 January 2018 (05.01.2018)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
62/447,851 18 January 2017 (18.01.2017) US
- (71) **Applicant:** ILLUMINIA, INC. [US/US]; 5200 Illumina Way, San Diego, California 92122 (US).
- (72) **Inventors:** WU, Kevin; 5200 Illumina Way, San Diego, California 92122 (US). ZHAO, Chen; 5200 Illumina Way, San Diego, California 92122 (US). CHUANG, Han-Yu; 5200 Illumina Way, San Diego, California 92122 (US). SO, Alex; 5200 Illumina Way, San Diego, California 92122 (US). TANNER, Stephen; 5200 Illumina Way, San Diego, California 92122 (US). GROSS, Stephen M.; 5200 Illumina Way, San Diego, California 92122 (US).

- (74) **Agent:** ZHOU, Feng 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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).

- (54) **Title:** METHODS AND SYSTEMS FOR GENERATION AND ERROR-CORRECTION OF UNIQUE MOLECULAR INDEX SETS WITH HETEROGENEOUS MOLECULAR LENGTHS



- (57) **Abstract:** The disclosed embodiments concern methods, apparatus, systems and computer program products for determining sequences of interest using unique molecular index sequences that are uniquely associable with individual polynucleotide fragments, including sequences with low allele frequencies and long sequence length. In some implementations, the unique molecular index sequences include variable-length nonrandom sequences. In some implementations, the unique molecular index sequences are associated with the individual polynucleotide fragments based on alignment scores indicating similarity between the unique molecular index sequences and subsequences of sequence reads obtained from the individual polynucleotide fragments. System, apparatus, and computer program products are also provided for determining a sequence of interest implementing the methods disclosed.

Published:

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