

(51) International Patent Classification:
G16B 50/50 (2019.01) *H03M 7/30* (2006.01)(21) International Application Number:
PCT/US2021/058364(22) International Filing Date:
05 November 2021 (05.11.2021)

(25) Filing Language: English

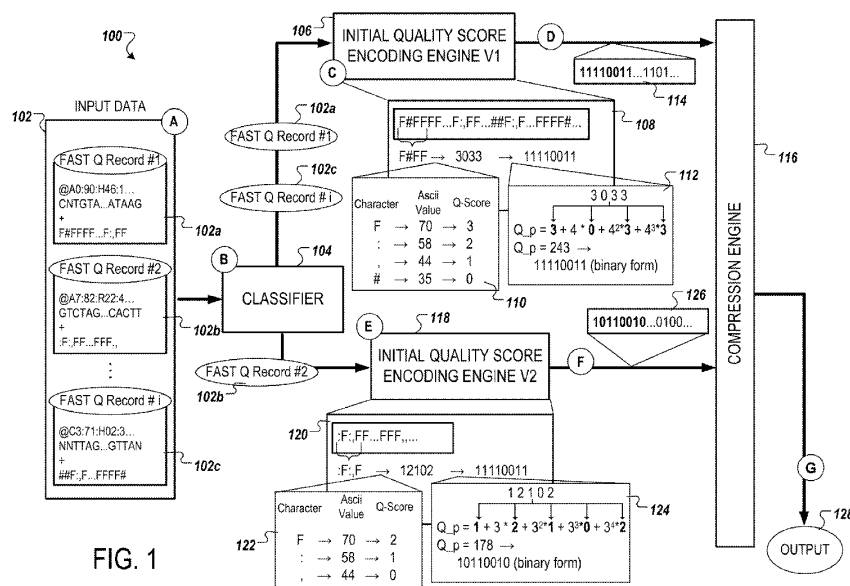
(26) Publication Language: English

(30) Priority Data:
63/110,308 05 November 2020 (05.11.2020) US(71) Applicant: **ILLUMINA, INC.** [US/US]; 5200 Illumina Way, San Diego, California 92122 (US).(72) Inventor: **RIZK, Guillaume Alexandre Pascal**; 5200 Illumina Way, San Diego, California 92122 (US).(74) Agent: **DARNO, Patrick** et al.; Fish & Richardson P.C., P. O. Box 1022, Minneapolis, Minnesota 55440-1022 (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, 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,

(54) Title: QUALITY SCORE COMPRESSION



(57) **Abstract:** Methods, systems, and computer programs for compressing nucleic acid sequence data. A method can include obtaining nucleic acid sequence data representing: (i) a read sequence, and (ii) a plurality of quality scores, determining whether the read sequence includes at least one N base, based on a determination that the read sequence does not include at least one N base, generating a first encoded data set by using a first encoding process to encode each of the quality scores of the read sequence using a base-(x minus 1) number, where x is an integer representing a number of different quality scores used by the nucleic acid sequencing device, and using a second encoding process to encode the first encoded data set, thereby compressing the data to be compressed.

[Continued on next page]

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))*

Published:

- *with international search report (Art. 21(3))*