



- (51) **International Patent Classification:**
C12Q 1/6809 (2018.01) *G06F 17/18* (2006.01)
C12Q 1/6827 (2018.01)
- (21) **International Application Number:**
PCT/SE2019/050299
- (22) **International Filing Date:**
02 April 2019 (02.04.2019)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
1850405-0 11 April 2018 (11.04.2018) SE
- (72) **Inventors; and**
(71) **Applicants:** YUDOVICH, David [SE/SE]; Kulgränden 13A, 22649 LUND (SE). LARSSON, Jonas [SE/SE]; Skyttelinjen 136, 226 49 LUND (SE).
- (74) **Agent:** ZACCO SWEDEN AB; Box 5581, 114 85 STOCKHOLM (SE).
- (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,

(54) **Title:** DETERMINATION OF FREQUENCY DISTRIBUTION OF NUCLEOTIDE SEQUENCE VARIANTS

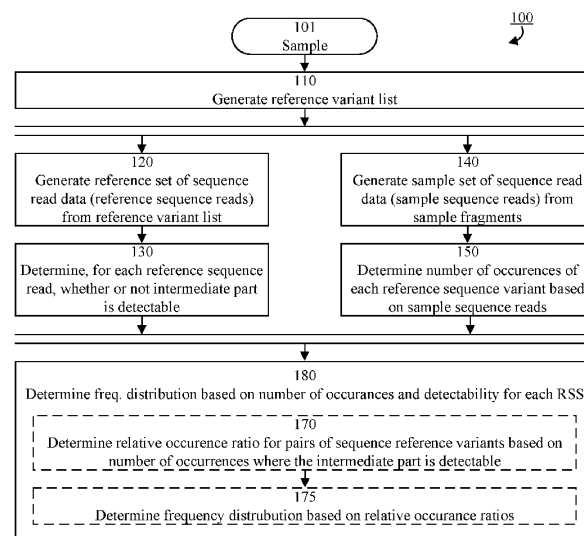


FIG. 1

(57) **Abstract:** Determination of a frequency distribution of nucleotide sequence variants - differing from a nucleotide reference sequence by at least one intermediate part - of a sample is disclosed. A reference set of sequence read data is generated by performing a plurality of reference sequence reads (each having a read start site, RSS) for each sequence reference variant of a reference variant list. It is determined (for each RSS and for the intermediate part of each sequence reference variant) whether the intermediate part is detectable by corresponding reference sequence read. A sample set of sequence read data is generated by performing sample sequence reads for the sample, and a total number of occurrences in the sample set is determined for each RSS and for the intermediate part of each sequence reference variant. The frequency distribution of the plurality of nucleotide sequence variants is determined based on the number of occurrences determined for an RSS where the intermediate part is detectable.

[Continued on next page]

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).

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i))*

Published:

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